

1988

Foundational Issues In The Representational Theory Of Mind

Mary Frances Egan

Follow this and additional works at: <https://ir.lib.uwo.ca/digitizedtheses>

Recommended Citation

Egan, Mary Frances, "Foundational Issues In The Representational Theory Of Mind" (1988). *Digitized Theses*. 1727.
<https://ir.lib.uwo.ca/digitizedtheses/1727>

This Dissertation is brought to you for free and open access by the Digitized Special Collections at Scholarship@Western. It has been accepted for inclusion in Digitized Theses by an authorized administrator of Scholarship@Western. For more information, please contact tadam@uwo.ca, wlsadmin@uwo.ca.



National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

AVIS

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

FOUNDATIONAL ISSUES IN THE REPRESENTATIONAL THEORY OF MIND

by

Mary Frances Egan

Department of Philosophy

Submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario
April 1988.

© Mary Frances Egan 1988

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-43288-8

Foundational Issues in the Representational Theory of Mind

ABSTRACT

A comprehensive theory of mind should satisfy two constraints or conditions: (1) it should provide an account of the proper construal of propositional attitudes, e.g. beliefs and desires, in folk psychological explanations of behaviour, and (2) it should explain how propositional attitudes can be causally efficacious in the production of behaviour. The representational theory of mind (hereafter the RTM) proposes to satisfy (1) by construing propositional attitude tokenings as formal processes defined over (semantically interpreted) symbol structures. The RTM proposes to satisfy (2) by respecting what has come to be known as the 'formality' condition, which requires that mental processes have access only to the formal (i.e. non-semantic) properties of the structures over which they are defined. According to the RTM, the mental representations which the theory postulates have their causal roles in virtue of their syntax.

In the thesis I evaluate the prospects for a theory of mind that (i) individuates mental states by reference to their contents and (ii) construes mental processes as syntactic processes. I first consider a proposal for explaining behaviour without recourse to semantically

interpreted states, arguing that purely syntactic theories will be unable to capture the wide range of behavioural regularities readily explained by content-based theories. I then examine and criticize a number of recent proposals, notably those of Fodor and Burge, for the individuation of mental states by their contents. I conclude that the arguments, on both sides, that allege to show that psychological theories must employ some antecedently specified notion of content are inconclusive. I then go on to criticize the central tenet of the RTM -- that syntactic descriptions, rather than semantic descriptions, capture the intrinsic causal properties of the mind -- arguing that this doctrine is based on the mistaken belief that propositional attitudes must be construed as relational states to satisfy the two adequacy conditions on a theory of mind. I conclude the thesis with a sketch of a non-relational construal of propositional attitudes which, I argue, seems at least as likely as the RTM to satisfy the aforementioned adequacy conditions.

Acknowledgements

I would like to thank my supervisor, Ausonio Marras, for his help and patience. I would also like to thank Ronald Clark, Dave Davies, Barry Loewer, Ernest Lepore, and Brian McLaughlin for many helpful discussions on issues in the philosophy of mind and psychology, and especially Robert Matthews for valuable criticism, advice, and encouragement.

TABLE OF CONTENTS

CERTIFICATE OF EXAMINATION	ii
ABSTRACT	iii
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vi
CHAPTER I -- INTRODUCTION: THE REPRESENTATIONAL THEORY OF MIND	1
Notes	23
CHAPTER II -- PSYCHOLOGY WITHOUT CONTENT: THE SYNTACTIC THEORY OF MIND	26
1. The Syntactic Theory	26
2. Do STM theories capture an interesting class of generalizations missed by content-based theories?	28
3. Do STM theories miss important generalizations captured by content-based theories?	36
Notes	44
CHAPTER III -- PSYCHOLOGY AND INDIVIDUALISM	47
1. An A Priori Argument for Individualism	52
2. Does Psychological Practice Support Individualism?	72
Notes	92
CHAPTER IV -- INTERMEZZO: THE FORMALITY CONDITION AGAIN	97
Notes	103
CHAPTER V -- PROBLEMS WITH THE REPRESENTATIONAL THEORY OF MIND ..	104
1. Two Kinds of Counterexamples	105
2. Adequacy Conditions on a Theory of Propositional Attitudes	111
3. Arguments for the Language of Thought	126
4. Intentional Realism Without Explicit Representation	145
Notes	151
BIBLIOGRAPHY	157
VITA	161

The author of this thesis has granted The University of Western Ontario a non-exclusive license to reproduce and distribute copies of this thesis to users of Western Libraries. Copyright remains with the author.

Electronic theses and dissertations available in The University of Western Ontario's institutional repository (Scholarship@Western) are solely for the purpose of private study and research. They may not be copied or reproduced, except as permitted by copyright laws, without written authority of the copyright owner. Any commercial use or publication is strictly prohibited.

The original copyright license attesting to these terms and signed by the author of this thesis may be found in the original print version of the thesis, held by Western Libraries.

The thesis approval page signed by the examining committee may also be found in the original print version of the thesis held in Western Libraries.

Please contact Western Libraries for further information:

E-mail: libadmin@uwo.ca

Telephone: (519) 661-2111 Ext. 84796

Web site: <http://www.lib.uwo.ca/>

Chapter 1

Introduction: The Representational Theory of Mind

It is a desideratum of any comprehensive theory of mind that it satisfy two conditions or constraints: (1) that it provide an account of the proper construal of propositional attitude ascriptions, i.e. ascriptions of beliefs and desires, in folk psychological explanations of behaviour, and (2) that it explain how propositional attitudes can be causally efficacious in the production of behaviour.¹

The rationale for condition (1) is the following: taken at face value, ascriptions of beliefs and desires to agents seem to ascribe to them causally efficacious internal states which mediate their behaviour. There is substantial disagreement, however, on how such ascriptions are to be construed. On the one hand, some have proposed outright elimination of propositional attitude discourse. Radical behaviourists² have argued that propositional attitudes are to be eliminated as excess metaphysical baggage in favour of the more materialistically respectable dispositions to behaviour, while eliminative materialists³ hold that propositional attitudes are the theoretical entities of an outmoded and essentially false empirical theory. On the other hand, there have been

various reductive proposals for retaining the folk psychological typology while reducing it to some less controversial basis. Logical behaviourists⁴ have held that propositional attitude ascriptions are to be understood as attributing to agents complexes of behavioural dispositions (e.g. "S believes that p" means that S is disposed to behave in some specified way), while reductive materialists⁵ have argued that propositional attitudes are contingently identical states of the central nervous system. In the face of such widespread disagreement about the proper role of appeals to beliefs and desires in explanations of behaviour, any theory of mind should specify precisely how ascriptions of propositional attitudes are to be construed and what commitments they involve.

Condition (2) requires, in effect, that a theory of mind explain how mental causation is possible. Folk psychology has nothing to say on this question: it simply assumes that the beliefs and desires it ascribes to agents are causally efficacious in producing their behaviour. To ask how this is accomplished is to ask for a solution to the mind/body problem. Any theory of mind, to merit the name, must explain how mental processes are related to physical processes, in particular, to behaviour. To be materialistically respectable, a theory of mind must do so

3

without postulating any mysterious, non-physical, causal powers.

The representational theory of mind (hereafter the RTM) construes mental processes as formal processes defined over (semantically interpreted) symbol structures. The RTM has been designed with the above conditions explicitly in mind. Indeed, its proponents see it as the only way to jointly satisfy the two constraints. In this introductory chapter I shall articulate the main theses of the RTM and explain how the theory is intended to satisfy the two conditions. In my exposition of the RTM I shall rely heavily on the views of its two chief proponents, Jerry Fodor and Zenon Pylyshyn.

The RTM proposes to satisfy condition (1) by construing propositional attitudes as relations between organisms and formulae in an internal language. More precisely, the RTM holds that for any organism O and attitude A toward the proposition p , there is a computational relation R and a mental representation MP such that MP means that p , and O has A to p if and only if O bears R to MP . The RTM further claims that mental processes are causal sequences of tokenings of mental representations.⁶ According to the RTM, then, propositional attitudes are individuated along two independent dimensions: by the content of a mental

representation and by a computational relation to that content.

Folk psychological explanations of behaviour are intentional explanations - they explain the behaviour of a cognitive agent by reference to her propositional attitudes, e.g. her belief that some state of affairs p obtains, her desire that some state of affairs q obtain, and so on. The explanations of behaviour licensed by the RTM preserve this feature of folk psychological explanations - they are also intentional. Like folk psychology, the RTM explains behaviour by reference to the contents of mental states. Proponents of the RTM argue that regularities important for the explanation of behaviour can only be captured in intentional terms.

Pylyshyn, for example, argues as follows:⁷ the folk psychological explanation of Mary's running out of the building appeals (inter alia) to Mary's belief that the building is on fire, i.e., to the way she represents the situation to herself. Such explanations will be more predictive than explanations that appeal only to physical features of the situation because the former will comprehend counterfactual cases which are beyond the explanatory scope of the latter. For example, suppose Mary in fact comes to have the belief that the building is on fire because she smells smoke. Then any explanation of

Mary's running-out-of-the-building behaviour couched in purely physical terms will appeal either to features of the distal cause (the smoke) or to features of Mary's perceptual apparatus (her olfactory system). But Mary might have arrived at the belief that the building is on fire by a very different route. She might have been informed by telephone by a neighbour who sees flames shooting out a window. Or she might have arrived at the (in this case mistaken) belief that the building is on fire by hearing a malfunctioning fire alarm. The crucial point is that these diverse circumstances need have nothing physical in common, so important regularities will be missed by any explanation of Mary's behaviour which appeals only to physical features of the situation. Mary's behaviour depends crucially on how she represents the situation to herself, that is, on what she believes (among other things, that the building is on fire), and on what she desires (not to be burned), so only a theory which appeals to the contents of Mary's mental states can capture and explain the relevant behavioural regularities.

The RTM, accordingly, postulates contentful states and appeals to such states in its explanations of behaviour. In so doing, proponents of the RTM hope to systematize and make rigorous the generalizations of folk psychology. Indeed, Fodor believes that cognitive

psychology, developed along the lines of the RTM, will "vindicate" folk psychology in the sense that mature theory will underwrite the intentional realism implicit in folk psychological explanations of behaviour. A theory is intentionally realist just in case it postulates states (entities, events) that satisfy the following conditions: (1) they are semantically evaluable, (2) they play a causal role in the production of behaviour, and (3) the generalizations of commonsense psychology are largely true of them.⁸ The RTM, according to its proponents, postulates such states.

It remains for the RTM to satisfy the second condition on a theory of mind - that it explain how mental causation is possible. The RTM proposes to do this by construing mental processes as computational processes. Pylyshyn claims that mental processes ought to be construed as computational processes because "computation is the only worked-out view of process that is both compatible with a materialist view of how a process is realized and that attributes the behaviour of the process to the operation of rules upon representations."⁹

The relevant notion of computation is explicated by Fodor as follows:

~~Every~~ computational device is a complex system which changes physical state in some way

7

determined by physical laws. It is feasible to think of such a system as a computer just insofar as it is possible to devise some mapping which pairs physical states of the device with formulae in a computing language in such fashion as to preserve desired semantic relations among the formulae. For example, we may assign physical states of the machine to sentences of the language in such a way that if $S_1 \dots S_n$ are machine states, and if $F_1 \dots F_{n-1}, F_n$ are the sentences paired with $S_1 \dots S_{n-1}, S_n$ respectively, then the physical constitution of the machine is such that it will actually run through that sequence of states only if $F_1 \dots F_{n-1}$ constitutes a proof of F_n . Patently, there are indefinitely many ways of pairing states of the machine with formulae in a language which will preserve this sort of relation, which is to say that the decipherment of the machine code exhibits indeterminacy of translation. Patently, there are indefinitely many ways of assigning formulae to machine states which do not preserve such relations among the formulae: only, in such assignments, we cannot interpret the machine's changes of state as proofs. 10

Fodor thus construes computation as mechanical theorem-proving, and the computational theory of mind, accordingly, interprets the sequence of physical states causally responsible for behaviour as a proof which has a description of the behaviour as its conclusion.

The idea is that, in the case of organisms as in the case of real computers, if we get the right way of assigning formulae to the states it will be feasible to interpret the sequence of events that causes the output as a computational derivation of the output. In short, the organic events which we accept as implicated in the etiology of behaviour will turn out to have two theoretically relevant descriptions if things turn out right: a

physical description by virtue of which they fall under causal laws and a psychological description by virtue of which they constitute steps in the computation from the stimulus to the response.

11

The following obvious worry arises with Fodor's notion of computation: if it is appropriate to construe the etiology of an organism's behaviour as the operation of rules upon representations, why should we not give a similar account of the behaviour of other physical systems, e.g. why not say that the planets follow Kepler's laws in their orbits around the sun? After all, we can always provide a computational description of a physical system by specifying a realization function (f_R), a 1-1 mapping which takes physical states of the system into a set of formulae in such a way that causal sequences of physical states can be interpreted as symbolic transformations. Fodor's response to this worry is that organisms, unlike the planets, are representation-using systems inasmuch as "a representation of the rules they follow constitutes one of the causal determinants of their behaviour".¹² So, whereas the planets can be said to compute the laws of planetary motion, on the weak notion of computation that Fodor provides in the passage quoted above, we are not inclined to attribute to them a system of representations whereby they represent the laws to

9

themselves; we can explain their behaviour without such attributions. The RTM, then, claims not merely that cognitive systems are computational systems, it claims that they are computational systems of a specific sort, viz. representation-using systems.

The RTM proposes to satisfy the requirement that a theory of mind explain how mental states can be causally efficacious in the production of behaviour by postulating physically instantiated internal representations as the causes of behaviour. Additionally, the RTM respects what Fodor calls the formality condition. In "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology", he claims

...the computational theory of mind requires that two thoughts can be distinct in content only if they can be identified with relations to formally distinct representations. More generally, fix the subject and the relation, and then mental states can be (type) distinct only if the representations, which constitute their objects are formally distinct. 13

The formality condition, as explicated by Fodor, requires that differences in the content of two representations are reflected in differences in their forms. In other words, mental processes have access only to formal properties of representations.

Although it is clear that formal properties are not

semantic properties, Fodor is otherwise vague on what counts as a formal property. Formal operations, he claims, "apply in terms of the, as it were, 'shapes' of the objects in their domains."¹⁴ Formal properties are sometimes identified with syntactic properties, and the formality condition accordingly interpreted as holding that mental states/processes are distinct in content only if they are syntactically distinct (or, alternatively, that content supervenes on syntax). But this is not strictly correct. The formality condition itself does not require that mental states have any syntactic properties at all, or that mental operations involve any syntactic operations. A syntactic operation is a type of formal operation, according to the above criteria, but it is simply one type of formal operation. Many physical operations (protein synthesis, for example) apply in virtue of the shape of the objects in their domains and so would count as formal operations on the above criteria. What constraints the formality does put on psychological theories will be discussed in chapter 4.

The RTM, however, does claim that mental states/processes are both formal and symbolic. They are symbolic because they are defined over symbol structures, and they are formal, according to the RTM, because mental operations apply in virtue of the syntax of these symbol

structures. Thus, the RTM honours the formality condition by claiming that mental operations are syntactic operations, a species of formal operations. It is perhaps natural, then, for computationalists to identify the formal properties of a mental state with its syntactic properties. The RTM explicitly claims that it is by virtue of the syntactic features of mental representations, realized in causally efficacious physical structures, that these representations produce behaviour. In other words, mental representations have their causal roles in virtue of their syntactic properties. Therefore, the RTM promises to satisfy the second constraint on a theory of mind - that it explain how propositional attitudes can be causally efficacious in the production of behaviour - by treating mental processes as physically realized syntactic processes.

I can now summarize the main theses of the RTM. The RTM claims that (1) propositional attitudes are relational states; (2) among the relata are mental representations, which have both semantic and syntactic properties; (3) mental representations have their causal roles in virtue of their syntactic properties (4) mental processes are computational processes; (5) the generalizations of psychological theory advert to the content of mental states. 15

The formality condition, according to Fodor, requires a sort of methodological solipsism:¹⁶

If mental processes are formal, then they have access only to the formal properties of such representations of the environment as the senses provide [sic]. Hence, they have no access to the semantic properties of such representations, including the property of being true, of having referents, or, indeed, the property of being representations of the environment.
17

Fodor contrasts psychological theory which respects the formality condition, and hence is methodologically solipsistic, not raising questions about truth and reference (i.e. about the way that thought relates to the world), with naturalistic psychology, which is concerned with the (presumably) causal relations between an organism's representations and the environment. He claims that much current psychological practice is methodologically solipsistic, particularly research involving machine simulation. He argues further that at least some part of cognitive psychology should be methodologically solipsistic, i.e. should respect the formality condition. He offers two independent arguments for the latter claim.

The first argument takes the following form:¹⁸ (1)

It is typically under an opaque construal that attributions of propositional attitudes to organisms function in explanations of their behaviour. (2) The

formality condition is "intimately involved" [Fodor's term] in the explanation of propositional attitudes opaquely construed; only theories which respect it will yield such explanations. (3) Hence, some part of psychology (i.e. that part concerned with the explanation of behaviour) should respect the formality condition.

In support of premise (1), Fodor points out that opaquely construed propositional attitude ascriptions¹⁹ are true in virtue of the way the agent represents the situation to himself. He argues further, from considering a number of examples similar to Pylyshyn's above, that the agent's representations of the situation function crucially in the production of his behaviour. Therefore, Fodor concludes, the opaque reading of propositional attitude ascriptions, not the transparent reading, allows us to predict and explain the behaviour of cognitive agents.

Premise (2) is more problematic. It is unclear what Fodor means by the claim that the formality condition is "intimately involved" in the explanation of propositional attitudes opaquely construed. He does say that "the notion of same mental state that we get from a theory which honours the formality condition is related to, but not identical to, the notion of same mental state that unconstructed intuition provides for opaque construals."²⁰

He shows by a series of examples that the conditions on opaque type identification are not strictly formal, but are partly semantic. Fodor's examples involve pronouns and demonstratives; they indicate that in such cases referential considerations play a role in the type-identification of beliefs, even where the relevant belief ascriptions are opaquely construed. Some examples show that formal identity of mental representations is not necessary for type identity of opaquely construed mental states (e.g. Both you and I think that I'm sick. I think: "I'm sick", while you think "she's sick". Our thoughts are opaquely type-identical, but formally distinct.) Others show that it is not sufficient (e.g. Alfred thinks of Misha, "he feels faint", while Misha thinks of Sam, "he feels faint". Their thoughts are opaquely type-distinct, although formally identical.) The insufficiency of formal identity for content identity is especially troubling, given that the formality condition requires that differences in the content of representations are reflected in differences in their forms. So it is not strictly true that a psychology that respects the formality condition studies propositional attitudes opaquely taxonomized. Yet, according to premise (1), the opaque construal of propositional attitudes is the appropriate one for understanding folk psychological

explanations of behaviour. In effect, then, a theory that respects the formality condition will require a different notion of same mental state than the folk psychological notion.

Proponents of the RTM have firmly believed that their theory will both explain how mental causation is possible (i.e. solve the mind/body problem) and "vindicate" folk psychology. It is not obvious that the RTM can do both. It proposes to do the former by respecting the formality condition, but, in so doing, it seems to require a different taxonomy of mental states than that employed by folk psychology, at least if we assume, as Fodor does, that folk psychology type-identifies mental states opaquely. Since the core idea of folk psychology is that cognitive agents act out of their beliefs and desires, and since beliefs and desires are type-individuated by their contents, a theory which taxonomizes mental states by their contents would seem to have the best hope of systematizing folk psychology's generalizations. The question, then, is whether a psychological theory can taxonomize mental states by their contents and respect the formality condition. This issue will be addressed in chapter 3.

Fodor's argument for methodological solipsism, then, is inconclusive: theories that respect the formality

condition do not seem to yield the appropriate notion of content.

Fodor has a second argument for the claim that at least some part of cognitive psychology should be methodologically solipsistic. He is skeptical about the possibility of a naturalistic psychology, that is, a theory of the organism/environment relations that fix the semantic properties of mental states. He argues that "...characterizing the objects of thought is methodologically prior to characterizing the causal chains that link thoughts to their objects. But the theory which characterizes the objects of thought is the theory of everything; it's all of science."²¹ Thus, he claims, a naturalistic psychology must await the completion of the rest of science. A methodologically solipsistic psychology, therefore, is all we can practically hope to get. Fodor's reasoning, however, is unconvincing. He claims, correctly, that "a naturalistic psychology would attempt to specify environmental objects in a vocabulary such that environmental/organism relations are law instantiating when so described,"²² but he provides no argument for the claim that such descriptions must await the completion of other sciences (presumably, he has in mind physics and chemistry). What the naturalistic psychologist needs are descriptions of the objects of

thought such that, under such descriptions, environment/organism relations instantiate laws of her own theory, not laws of physics or chemistry. The appropriate vocabulary may turn out to be a vocabulary referring to observable properties. Consider frog/fly interactions: the relevant generalizations do not depend on a theory of flies, since the frog is cueing on rather simple observable properties of flies, as is shown by the fact that we can elicit the same behaviour using suitably moving fly-sized BBs.²³ The important point is that the generalizations describing such interactions, which refer to observable properties shared by flies and BBs, are not likely to instantiate any laws of physics or chemistry, but may well instantiate laws of a naturalistic psychological theory. Since Fodor has given no reason to think that these laws must await the completion of the other sciences, he has provided no grounds for skepticism about the prospects of naturalistic psychology.

There have been various criticisms of the view that the RTM can and should be methodologically solipsistic in framing its explanations of behaviour. Stephen Stich, in his book From Folk Psychology to Cognitive Science,²⁴ has argued that what he calls the "strong version" of the Representational Theory of Mind violates the principle of

methodological solipsism. (Stich's "strong" RTM thesis claims that "nomological generalizations which describe the interactions among mental states apply to them in virtue of their contents."²⁵ The RTM, as I have articulated it here, taxonomizes mental states by their contents, and so counts as an instance of the "strong" thesis.) Stich construes methodological solipsism as the doctrine that "cognitive psychology ought to restrict itself to postulating formal operations on mental states. It ought not to postulate processes which apply to mental states in virtue of their semantic properties." ²⁶

According to Stich,

... methodological solipsism...entails the rejection of the strong RTM. For on any plausible account of what counts as semantic, the theorist who couches his generalizations (his account of mental processes) in terms of the content sentences used to characterize mental states is surely postulating mental operations whose specification requires reference to semantic properties of these states. 27

Stich seems to have offered a decisive argument against the possibility that a psychological theory can both honour the formality condition and taxonomize mental states by their contents (and hence frame its generalizations in terms of content). However, Stich is confused about the import of both methodological solipsism

and the formality condition for the psychological theorist. Fodor proposes the formality condition (mental states are content-distinct only if they are formally distinct) as a constraint that must be satisfied by any psychological theory which aspires to be methodologically solipsistic. However, Stich claims that the formality condition requires "that semantic properties of mental states play no role in the specification of psychological generalizations."²⁸ This is simply false. The formality condition, as articulated by Fodor and Pylyshyn, requires that every semantic difference recognized by a theory entails a formal difference. Thus it requires that content supervene on form, that no psychological processes depend upon the semantic properties of mental states, but it entails nothing about how the generalizations of psychological theories are to be framed. The psychological theorist is free to advert to the semantic properties of mental states in her explanations of behaviour, as long as she explains in virtue of what formal property a mental state has the content that it is claimed by the theory to have. Therefore, Stich is simply wrong in claiming that methodological solipsism and the formality condition require the rejection of psychological theories whose generalizations advert to content.

In subsequent chapters I discuss a number of foundational issues in the representational theory of mind. The first set of issues centres around the role of content in the theory, the second set concerns the role of syntax. My overall conclusions are that the problems alleged to affect content-based theories are not insurmountable. Appeal to content does seem to be indispensable for a theory which purports to explain behaviour, but the RTM's claims that a syntactic level of description is indispensable are unsupported.

In chapter two I consider an alternative proposal, the syntactic theory of mind, which attempts to explain behaviour without recourse to semantically interpreted states. Stich claims that syntactic theories can capture a wider range of generalizations about cognitive processes than representationalist theories, but this claim, I argue, is unfounded. Not only is it false that syntactic theories can capture psychological generalizations that content-based theories cannot, but a large class of behavioural regularities, readily explained by content-based theories, appears to be beyond their explanatory reach.

In chapter three I address the question raised in this introduction: what notion of content is available to the RTM, and will the available notion enable it to

capture commonsense generalizations about behaviour? This question has been framed as a dispute about individualism, the question of whether content must supervene on non-intentionally specifiable properties of the individual subject. I criticize an a priori argument of Fodor's to the effect that a scientific psychology must be individualistic. I then consider Tyler Burge's arguments on the other side of the dispute, concluding that attempts to show that psychology is not individualistic are inconclusive. The RTM has been criticized on the grounds that it needs a notion of "narrow content" and can't have one. I argue that this criticism is unfounded. My conclusions in this chapter are that a psychological theory may choose whatever notion of content is appropriate to its explanatory goals.

Chapter four is a short discussion of the formality condition and the role it plays in the RTM. I conclude that it places less severe constraints on a theory of mind than proponents of the RTM have claimed.

Chapter five is a critique of the main tenet of the RTM -- that propositional attitudes are to be analysed as relations between organisms and formulae in an internal language. I criticize a number of arguments alleged to support this thesis, and in the process I sketch the outlines of a view which does not treat propositional

attitudes as relational states. I conclude by arguing that the alternative seems to have at least as good a chance of vindicating folk psychology as the RTM, yet it does so without the RTM's commitment to a language of thought.

Notes

1. To the extent that a theory of mind does not take attitude ascriptions to attribute causally efficacious states to an agent, the theory owes us an account of the role of attitude ascriptions in psychological discourse.
2. See e.g. Skinner, B.F. "Behaviourism at Fifty" in T.W.Wann, ed., Behaviourism and Phenomenology: Contrasting Bases for Modern Psychology, Chicago: University of Chicago Press, 1964.
3. See Churchland, P.M. Scientific Realism and the Plasticity of Mind, Cambridge: Cambridge University Press, 1979.
4. See Ryle, Gilbert. The Concept of Mind, London: Hutchinson, 1949.
5. See e.g. Armstrong, David. A Materialist Theory of the Mind. London: Routledge and Kegan Paul, 1968.
6. This is Fodor's formulation in Psychosemantics: The Problem of Meaning in the Philosophy of Mind, Cambridge, Mass.: MIT Press, 1987, p.17.
7. Pylyshyn, Z.W. "Cognitive Representation and the Process-Architecture Distinction," The Behavioural and Brain Sciences, 3, (1980), p.161.
8. Psychosemantics, p.10.
9. Pylyshyn, Z.W. "Computation and Cognition: Issues in

- the Foundations of Cognitive Science," The Behavioural and Brain Sciences, 3, (1980), p.113.
10. Fodor, J.A. The Language of Thought, New York: Thomas Y. Crowell Co., 1975, p.73.
 11. Ibid., pp.73-74.
 12. Ibid., p.74 (Fodor's emphasis)
 13. Fodor, J.A. "Methodological Solipsism Considered as a Research Strategy in the Cognitive Psychology," The Behavioural and Brain Sciences, 3, (1980), p.64.
 14. Ibid., p.64.
 15. Fodor, J.A. "Introduction," RePresentations. Cambridge, Mass.: MIT Press, 1981, p.26.
 16. The principle of methodological solipsism as articulated by Putnam in "The Meaning of Meaning" holds that "no psychological state, properly so-called, presupposes the existence of any individual other than the subject to whom that state is ascribed". (Collected Works, Vol.1, Cambridge: Cambridge University Press, 1975, p. 220)
 17. "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology," p.65.
 18. Ibid., p.66.
 19. i.e. the standard tests for transparency - substitution of identicals and existential generalization - typically fail.

20. Ibid., p.67.
21. Ibid., p.70.
22. Ibid., p.70.
23. I owe this example to Robert Matthews.
24. Stich, S.P. From Folk Psychology to Cognitive Science: The Case Against Belief, Cambridge, Mass: MIT Press, 1983.
25. Ibid., p.129.
26. Ibid., p.162 (emphasis in the original)
27. Ibid., p.162.
28. Ibid., p.162.

Chapter 2

Psychology Without Content: The Syntactic Theory of Mind

1. The Syntactic Theory

In his book From Folk Psychology to Cognitive Science¹ Stephen Stich develops an alternative to the representational theory of mind (RTM) defended by Fodor and Pylyshyn. The Syntactic Theory of Mind (hereafter STM) construes cognitive states as uninterpreted syntactic objects:

...cognitive states whose interaction is (in part) responsible for behaviour can be systematically mapped to abstract syntactic objects in such a way that causal interactions among cognitive states, as well as causal links with stimuli and behavioral events, can be described in terms of the syntactic properties and relations of the abstract objects to which the cognitive states are mapped. More briefly, the idea is that causal relations among cognitive states mirror formal relations among syntactic objects. If this is right, then it will be natural to view cognitive state tokens as tokens of abstract syntactic objects. 2

According to Stich, a cognitive theory which instantiates the STM is to be articulated in three parts: (i) a specification of a finite set of primitive syntactic objects and a set of formation rules which can generate (infinitely many) complex formulae; (ii) the hypothesis

that for each model of the theory there exists a set of state types (presumably neurological state types) whose tokens are causally efficacious in the production of behaviour, and further, that there is a mapping from these state types to syntactic objects in the class specified in (i); and (iii) a specification of the theory's lawlike generalizations. Such theories may plausibly be construed as constitutive of an abstract neurology, inasmuch as equivalence classes of neurological states are mapped into a space of syntactic objects; the equivalence classes are defined by the causal role of the states in the production of behaviour.

The generalizations of STM theories, which characterize the causal relations among neurological states, are said to be "specified indirectly via the formal relations among the syntactic objects to which the neurological state types are mapped."³ Stich gives the following as a typical example of such a generalization:

For all subjects S, and all wffs A and B, if S has a B-state mapped to $A \rightarrow B$ and if S comes to have a B-state mapped to A, then S will come to have a B-state mapped to B.⁴

STM theories, Stich argues, are to be preferred to theories whose generalizations advert to the content of mental states (i.e. representationalist, or RTM theories), because they can (i) "do justice to" all of a content-

based theory's generalizations, and (ii) capture additional generalizations which are beyond the reach of content-based theories. I shall challenge both claims. I shall begin with the latter claim, arguing that generalizations beyond the explanatory reach of content-based theories are likely to be beyond the reach of STM theories as well. I shall then turn to Stich's first claim, arguing that STM theories fail to capture behavioural regularities readily explained by content-based theories.

2. Do STM Theories Capture an Interesting class of generalizations missed by content-based theories?

A substantial part of Stich's book is devoted to establishing that the ascription of belief in folk psychology is both observer-relative, and irremediably vague. The ascription of belief is observer-relative, Stich argues, inasmuch as a subject's belief is identified as the belief that p just in case it is content similar to the belief which underlies our own normal assertion of "p". It is vague inasmuch as (i) observers are said to judge content similarity along at least three distinct dimensions of similarity and (ii) pragmatic factors (viz., context) determine which of these dimensions plays a

decisive role in a given judgment of content similarity. The three dimensions that Stich identifies are "causal-pattern similarity" (patterns of causal interaction with stimuli, behaviour, and other mental states; elsewhere called "narrow causal role"), ideological similarity (similarity of doxastic surroundings), and referential similarity.

The observer-relativity of folk psychological ascriptions of belief is alleged to create a special problem for content-based theories, since for each dimension of similarity there are said to be subjects sufficiently different from ourselves as to make it impossible to find a content sentence to describe their beliefs. Since generalizations stated over content sentences will therefore have indeterminate application in these cases, so-called "exotic" subjects will be beyond the explanatory reach of content-based theories.

Syntactic theories, by contrast, will allegedly have no special difficulty characterizing the mental states of exotic subjects, as they characterize mental states not by content sentences but by the syntactic objects to which the states are mapped. The latter are selected by the theorist with an eye to giving the simplest and most powerful account of the causal links among stimuli, mental states, and behavior, and without any concern for similarities and dissimilarities between the subject and

the theorist." ⁵

Stich assembles what I take to be a compelling case for the claim that belief ascription in folk psychology is vague and observer-relative. The relevant question, however, is whether ascriptions of content in a developed, scientific psychology will, of necessity, be equally vague and observer-relative. Stich provides no argument that they will. A scientific psychology might plausibly be expected to reduce the vagueness inherent in our everyday folk notion of content. After all, the folk notion serves many purposes besides the scientific ones of prediction and explanation. Thus, while ascriptions of content in folk psychology may indeed depend pragmatically on three distinct notions of content similarity, it is not obvious that ascriptions of content in a scientific psychology need do so. In a series of recent papers, for example, Jerry Fodor has attempted to articulate a more austere notion of content than that ascribed in folk psychology, viz. so-called narrow content ⁶. The narrow content of a belief is said to supervene on physical and/or functional states of the believer, thus referential similarity plays no role in judgments of narrow content similarity. If a notion of narrow content could be articulated and put to use in an RTM theory, ⁷ then subjects would not be classed as exotics whose mental states defy subsumption under generalizations defined over contents merely because their

beliefs were referentially different from our own.

But whatever the prospects for a notion of content that eliminates the vagueness and observer-relativity that seemingly afflicts folk psychological theories, STM theories seem no better able than content-based theories to characterize adequately the mental states of exotic subjects. Consider Stich's case of Mrs. T, an elderly subject whose memory has deteriorated as a result of degenerative brain disease to the point where her remaining beliefs are no longer ideologically similar to those that we would express using the same content sentences. (Two beliefs are ideologically similar if they are embedded in similar belief networks.) Before the onset of her illness, Mrs. T. believed that President McKinley was assassinated. After her memory has deteriorated, she is still disposed to answer "McKinley was assassinated" when asked "What happened to McKinley?", but when asked "Is McKinley dead?" she claims not to know. Clearly, we are disinclined to ascribe to Mrs. T. now the belief that she once had, viz. that McKinley was assassinated. For her current belief, whatever it is, is not the belief that would underlie our own sincere utterance of "McKinley was assassinated", because it fails to exhibit the appropriate connections to other beliefs. Indeed, we are at a loss to say what Mrs. T. believes now, because we have no content sentence available to

characterize her mental state. The problem for content-based theories becomes more acute if we assume that certain of Mrs. T.'s inferential skills are intact, e.g., that if told "If McKinley was assassinated then he is buried in Ohio" she will still respond "Then McKinley is buried in Ohio". For now there would seem to be a generalization regarding her behaviour, namely the STM generalization cited earlier, that a content-based theory is unable to capture because it has no way of characterizing the mental state which the generalization subsumes.

Stich claims that an STM theory will have no trouble characterizing Mrs. T.'s current mental state and subsuming it under the same generalization that described her earlier behaviour. I believe this claim to be false. Consider first what the claim that a theory can characterize a mental state amounts to. Clearly, the theory must do more than simply assign a name to the state; the theory must individuate the state, that is, it must be able to identify token occurrences of the state over time and across subjects in such a way that the state can be subsumed under lawlike generalizations that predict and explain the behaviour of subjects with a good measure of generality. This is precisely what content-based theories seem unable to do with respect to ideologically exotic subjects. STM theories, Stich argues, can so

characterize an exotic subject's mental states:

For a syntactic theory, however, ideological similarity poses no problem, since the characterization of a B-state does not depend on the other B-states that a subject happens to have. A B-state will count as a token of a wff if its potential causal links fit the pattern detailed in the theorist's generalizations, regardless of the further B-states the subject may have or lack. Consider, for example, the case of Mrs. T. If we assume that before the onset of her disease the B-state which commonly caused her to say "McKinley was assassinated" obeyed generalizations like [the sample STM generalization cited above], then if the illness simply destroys B-states (or erases mental tokens) without affecting the causal potential of the tokens which remain, the very same generalizations will be true of her after the illness has become quite severe. 80

While it seems possible that a mental token may simply be erased without affecting the individuation of the remaining states, this is not the correct description of what has happened to Mrs. T. The causal potential of Mrs. T's mental states have clearly changed: the mental state underlying the younger Mrs. T's utterance of "McKinley was assassinated" was potentially connected to behaviour and other mental states in countless ways; the mental state underlying the senile Mrs. T's utterance of "McKinley was assassinated", however, is causally inert, or nearly so.⁹ The latter state can no longer play the role in inference and belief fixation that characterized the former state; counterfactuals true of the earlier

state are not true of the latter. It is now false, for example, to say of Mrs. T that she would respond with "McKinley was assassinated" if asked "How did McKinley die?"

The problem for the syntactic theorist is not simply that because the causal potential of the mental state underlying Mrs. T's utterance has changed, syntactic generalizations such as the one cited by Stich that were true of Mrs. T before the onset of her illness will fail to be true of her now. The syntactic theorist cannot even individuate Mrs. T's current mental state as a token of some general state type shared by normal subjects (including Mrs. T before the onset of her illness). All the syntactic theorist can do is designate her state as a token of a new state type virtually unique to Mrs. T. Stich embraces this expediency with alacrity. According to Stich,

...neither causal nor ideological distance poses any special problems for an STM theorist. To handle subjects whose basic cognitive processes differ from our own, the syntactic theorist may specify a distinct set of wffs (a different 'mental language') and a distinct set of generalizations exploiting the syntactic structure of these wffs.

Consider once again the case of Mrs. T. She is, I have argued, a causal-pattern different subject. After her illness has become acute, a syntactic theorist may

simply construct a new syntactic theory and subsume the senile Mrs. T.'s mental states under the new theory's generalizations. The difficulty with Stich's proposal is precisely that the wffs assigned to Mrs. T.'s mental states by the new theory don't characterize the mental states of any other subjects, normal or exotic. The syntactic theorist can claim to have characterized Mrs. T.'s mental states in only a Pickwickian sense: he has succeeded in designating them as tokens of certain syntactic types, but they are types that lack all generality. Such "generalizations" as the syntactic theorist can construct for Mrs. T. lack all generality: they don't apply across subjects, even across most exotic subjects. They don't even apply to Mrs. T across time, since the causal potential of her mental states is continually changing. They apply only to Mrs. T as she is now.

This is individual psychology with a vengeance. The STM "handles" exotic subjects only insofar as it is willing to abandon our interest in both comparative and developmental psychology. At the very least Stich's proposal to construct a new theory for exotic subjects would make the comparison of exotic subjects and ourselves, or the comparison of different exotics, or the comparison of maturing or aging subjects over time, impossible. We have a strong intuition that our cognitive

processes have a lot in common with many subjects who count as exotic, for example, young children and animals. When we ascribe to a dog the belief that the squirrel ran up the tree, we attribute to it a mental state something like our own mental state when we believe that the squirrel ran up the tree. We want a theory that not only makes inter-species comparisons possible, but also applies diachronically to members of our own species.¹¹

In summary, Stich may be correct in arguing that content theories have a serious problem characterizing the mental states of exotic subjects; however, syntactic theories suffer the same problem. The wffs assigned to the senile Mrs. T.'s mental states characterize them only in a Pickwickian sense - the mental state types they pick out are virtually unique to Mrs. T. in her present condition. Stich's proposal to construct new theories to describe the behaviour of exotic subjects is a desperate measure - it abandons the project of a developmental or comparative psychology.

3. Do STM theories miss important generalizations captured by content-based theories?

However poorly content-based theories may fare at characterizing the mental states of exotic subjects, they

have been remarkably successful at predicting and explaining the behaviour of so-called normal subjects. There remains the question of whether STM theories can duplicate the success of folk psychology in the domain of normal subjects. Both Fodor and Pylyshyn have argued that regularities important for the explanation of behaviour can only be captured by generalizations which advert to the content of mental states.¹² These generalizations, it is claimed, are beyond the reach of STM theories.

Pylyshyn argues as follows:

It simply will not do as an explanation of, say, why Mary came running out of the smoke-filled building, to say that there was a certain sequence of expressions computed in her mind according to certain expression-transforming rules. However true that might be, it fails on a number of counts to provide an explanation of Mary's behaviour. It does not show why or how this behaviour is related to very similar behaviour she would exhibit as a consequence of receiving a phone call in which she heard the utterance "the building is on fire!", or as a consequence of hearing the fire alarm or smelling smoke, or in fact following any event interpretable (given the appropriate beliefs) as generally entailing that the building was on fire. 13

According to Pylyshyn, the only feature common to the diverse circumstances that would produce running-out-of-the-building behaviour on Mary's part is that they give rise to an internal state interpretable as a belief that the building is on fire. The relevant generalization, therefore, can only be captured by appeal to the content.

of Mary's internal states. Syntactic theories, of course, eschew appeals to content, and so, Pylyshyn claims, they will miss a generalization readily captured by content-based theories.

Stich's strategy in response to Pylyshyn's argument is to sketch the explanations that a content theory might give for Mary's behaviour and then construct parallel explanations in the STM mold. In version 1 of the content story Mary inhales smoke and is caused to believe that the building is on fire. This belief interacts with the long-standing conditional desire to leave a building if it is on fire to produce the desire to leave the building, which in turn interacts with the belief that if one runs out the door one will leave the building to produce the desire to run out the door. In version 2 of the content story Mary comes to believe that the building is on fire as a result of picking up the phone and hearing "The building is on fire!". From this point the story joins version 1.

Consider Stich's parallel STM explanations. In version 1 of the STM story Mary inhales smoke and is caused to have the B-state F (corresponding to the belief that the building is on fire¹⁴) by an indirect and complex causal process: inhaling the smoke causes her to have the B-state I, (corresponding to the belief that she is inhaling smoke) which interacts with the long-standing

B-state $I \rightarrow N$ (corresponding to the belief that if one is inhaling smoke then there is a fire nearby) which produces in her the B-state N , which together with the long-standing B-state $N \rightarrow F$ results in her having the B-state F (corresponding to the belief that the building is on fire). The B-state F interacts with the long-standing conditional D-state $F \rightarrow L$ (corresponding to the desire to leave the building if the building is on fire) to produce the D-state L , which in turn interacts with the B-state $R \rightarrow L$ (corresponding to the belief that if one runs out the door, then one will leave the building) to produce the D-state R , which eventuates in Mary's running out the door. In version 2 of the STM story, Mary's hearing "The building is on fire!" on the telephone causes her to have the B-state H (corresponding to the belief that she is hearing an utterance of "The building is on fire!") which leads her, via an indirect and complex causal chain, to have the B-state F , and from this point the story joins version 1.

What is the significance of the fact that parallel STM explanations can be given? Stich says

What the various versions have in common, according to the content-based strong RTM explanation, is that they all lead Mary to believe that the building is on fire, and this belief plays an essential role in the etiology of her fleeing behaviour. On the purely syntactic explanation, there is a prima facie perfect parallel. What the various versions

have in common is that they all lead Mary to have the B-state F, and this B-state plays an essential role in her fleeing behavior. 15

On both versions of the STM story, Mary's fleeing behaviour is caused by (among other things) her being in the B-state F. But this state arises, on the two versions, under diverse physical stimuli (distal and proximal). In fact, the causal sequences leading to the B-state F on the two versions of the story are completely different. What justification can the STM theorist provide for supposing that the two state tokens have anything syntactic in common if they don't play the same causal roles with respect to either stimuli or antecedent mental states? The only reason to identify the expressions designated by 'F' in the two versions is that they both mean "the building is on fire", but of course the STM theorist can't appeal to meanings in the individuation of mental states.

Stich repeatedly stresses the parallelism between content explanations and his own (STM) explanations, but the parallelism is not innocent. Underlying the STM accounts is the following assumption:

...each of the distinct content sentences used to characterize beliefs and desires in the strong RTM explanation corresponds to a distinct syntactic string....All that is being assumed is a token-token correspondence: each of Mary's beliefs and desires (i.e. each token) corresponds to a token of a syntactic

type, and the syntactic tokens are type distinct when and only when the corresponding belief and desire is accorded a distinct content sentence [my emphasis]. 16

It is unclear why Stich thinks that the assumption involves only a token-token correspondence. In fact, what is being assumed, as the underlined passage indicates, is a one-one mapping of semantic types onto syntactic types, that is, a type-correspondence between content and syntax. No argument is given for this very strong assumption, and it is, in fact, unlikely to be satisfied, given that types at the two levels of description are to be individuated on independent grounds: syntactic types by the causal roles of mental states in the production of behaviour, and semantic types by the various criteria that are involved in content ascription. Since the STM is plausibly construed as abstract neurology, satisfaction of the type-correspondence assumption would require, in effect, that contents map one-one onto equivalence classes of neurological states, such equivalence classes to be defined by the roles of these states in the etiology of behaviour. It is the burden of a large part of Stich's book that such a correspondence between the contents of mental states and their causal roles can't be effected; indeed if it could, then Stich's claim that content-based theories miss generalizations that can be captured by syntactic theories would be patently false, since both

types of theories would capture the same class of generalizations.

The type-correspondence assumption is unsupported, if not simply false. This leaves us with no way of type-identifying the internal states that give rise to Mary's fleeing behaviour under diverse circumstances except by appeal to their meaning. In particular, they cannot be type-identified by a theory which takes causal role to be criterial for the individuation of mental states, because the states are causally related to different stimuli and other internal states. But the STM explanations of Mary's behaviour hinge on the theory's ability to type-identify what Stich has called 'B-state F' in the two versions, otherwise a counterfactual supporting generalization is lost. The conclusion should be clear. Stich is faced with a dilemma: either the type-correspondence assumption is a reasonable assumption, and; consequently, semantic and syntactic generalizations will be co-extensive, thereby undermining Stich's arguments for preferring a syntactic theory to one which adverts to content; or the assumption is unjustified, and syntactic theories miss important generalizations about behaviour which can only be captured by content-based theories.

A final point about the parallelism between content and STM explanations: all the examples of STM generalizations in Stich's book are constructed to

parallel typical content-based generalizations. One must question whether a syntactic theory could stand on its own, whether it could do any genuine explanatory work. Perhaps the syntactic-based generalizations suggest themselves only where a content story has already been told. Nothing in Stich's book assuages these doubts. Not only does Stich fail to offer any empirical support for the STM, but the book contains not a single example of psychological research modelled on the STM. Stich relies exclusively on artificially constructed analogues of content-based explanations for his few examples of STM generalizations. In the absence of an explanatory practice that conforms to the STM pattern, one must seriously question Stich's claims that the STM provides an adequate, indeed preferred, foundation for psychological theorizing. At the very least, psychological theories constructed in the STM mold are unable to capture a wide range of folk psychological generalizations without exploiting the explanatory apparatus of folk psychology.

44

Notes

1. Stich, Stephen. From Folk Psychology to Cognitive Science: The Case Against Belief, Cambridge, Mass.: MIT Press, 1983.
2. Ibid., p.149.
3. Ibid., p.151.
4. This generalization is the syntactic analogue of the following typical content-based, or RTM, generalization:

For all subjects S, and all declarative sentences P and Q in our language, if S has a belief which can be attributed by a sentence of the form "S believes that if p then q" and if S comes to have a belief which can be attributed by a sentence of the form "S believes that p" then S will come to have a belief which can be attributed by a sentence of the form "S believes that q" where 'p' is replaced by P and 'q' is replaced by Q throughout.

5. Ibid., p.158.
6. His most recent proposal is in Psychosemantics, ch.2. See also "Cognitive Science and the Twin-Earth Problem", Notre Dame Journal of Formal Logic, 23, 1982;

- and White, S. "Partial Character and the Language of Thought," Pacific Philosophical Quarterly, 63, 1982.
7. Brian Loar, in "Social Content and Psychological Content" (unpublished manuscript), argues that conceptual role determines a notion of narrow content which already plays a prominent role in folk psychological explanation. See the discussion in chapter 3.
 8. Ibid., p.158.
 9. A similar point is made by Kim Sterelney in "Critical Notice: S.P.Stich, From Folk Psychology to Cognitive Science," Australasian Journal of Philosophy, 63, 1985 and by Ausonio Marras in "Critical Notice: Steven Stich's From Folk Psychology to Cognitive Science," Philosophy of Science, 63, 1987.
 10. Ibid., p.160.
 11. Such developmental theories as we now have suggest that a child undergoes maturational changes, that is, that the causal potential of the states change over time, yet there are good reasons for saying that the states persist through these changes. For example, work in psycholinguistics suggests that children learn grammar over a period of time. In developing learning theories, we want to be able to say that the rule that the child learns at three years old is the same rule that the adult knows, that is, we want to be able to

type-identify the state over time, across changes in causal potential. A series of syntactic theories, each individuating mental states by their causal potential, would not enable us to do this.

12. See Fodor, J.A. RePresentations, Cambridge, Mass.: MIT Press, 1981; Fodor, J.A. "Cognitive Science and the Twin-Earth Problem", Notre Dame Journal of Formal Logic, 23, 1982; Pylyshyn, Z.W. "Cognitive Representation and the Process-Architecture Distinction", Behavioural and Brain Sciences, 3, 1980; Pylyshyn, Z.W. Computation and Cognition: Towards a Foundation for Cognitive Science, Cambridge, Mass.: MIT Press, 1984.
13. Pylyshyn, Z.W. "Computation and Cognition: Issues in the Foundation of Cognitive Science," The Behavioural and Brain Sciences, 3, 1980, p.161.
14. These parenthetical remarks about what B- and D-states correspond to are inserted in Stich's text to make explicit the parallelism between STM and content-based explanations. They do not form any part of the STM explanations themselves. I shall have more to say on this point below.
15. Ibid., p.176.
16. Ibid., p.174.

Chapter 3

Psychology and Individualism

Jerry Fodor, in "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology,"¹ shows that the conditions on the type-identification of mental states in folk psychology are not strictly formal, but are partly semantic. In particular, Fodor shows that referential considerations play a role in the ascription of beliefs involving pronouns and demonstratives, because such terms are typically construed as referring, even when they occur in what are otherwise opaque contexts. For example, if Misha thinks of Sam "he feels faint", and Sam thinks of Misha "he feels faint", then the folk psychological theorist will ascribe different (de dicto) beliefs to the two subjects, despite the fact that the tokens "in their heads" are formally identical.

The formality condition requires that the contents of mental representations supervene on their formal properties.² If referential factors only play a role in mental state ascriptions involving pronouns and demonstratives, then the RTM theorist might simply deny that there are any genuinely indexical beliefs, claiming that the language of thought contains proper names where natural language belief ascriptions contain indexicals. The tokens in Misha's and Sam's heads when they each

believe that the other feels faint would, on this proposal, be formally distinct, and a computational psychology would be able to assign them different beliefs without violating the formality condition.

The problem Fodor points to in "Methodological Solipsism" seems to extend beyond indexicals, however, as is indicated by a now well-known set of examples. In Putnam's example,³ the inhabitants of twin-earth use the vocable "water" to refer not to H₂O, but to a superficially similar but chemically different liquid XYZ. Putnam uses twin-earth examples to argue that the meaning of natural kind terms is not in the head, but is partly determined by the causal history of the use of the term. If Putnam is right, then I and my twin-earth doppelganger mean different things when we utter the words "water is wet", and the beliefs underlying our assertion of these words have different contents, despite the fact that our mental states are formally identical... A psychological theory which type-identified our beliefs in respect of their contents would seemingly have to say that we have different beliefs, despite the fact that we are related to formally identical belief tokens.

In Tyler Burge's example,⁴ an individual, Burt, believes that he has arthritis in his thigh, not knowing that arthritis is only a disease of the joints. We are asked to compare Burt with a second subject, physically

and functionally identical to Burt, who lives in a community where the term "arthritis" is used to refer to a disease which affects not only the joints but the long bones of the legs as well. This second individual has a belief that he would express by the form of words "I have arthritis in my thigh". The intuition that Burge invites us to share is that Burt and his twin have different (de dicto) beliefs, that whatever twin-Burt believes, it is not a belief about arthritis. The two subjects are physically, functionally, and (up to the time when Burt learns that he is mistaken about arthritis) behaviourally identical, yet their thoughts have different contents, Burge argues, because meaning is partly determined by social factors outside the individual. If Burge is correct, then Putnam's dictum that meaning isn't in the head applies to all general terms, not just to natural kind terms.

In this chapter I shall consider what, if any, import the Putnam and Burge examples have for theory development in psychology. Burge takes the thought experiments to show that the content of propositional attitudes is not supervenient on non-intentionally specifiable properties or states of the individual subject. He concludes that a significant portion of psychology is, and will continue to be, non-individualistic. Burge defines individualism in psychology as follows:

According to individualism about the mind, the mental natures of all a person's or animal's mental states (and events) are such that there is no necessary or deep individuating relation between the individual's being in states of those kinds and the nature of the individual's physical or social environment. 5

As Burge understands the thesis, individualism about the mental holds that mental states are supervenient on non-intentionally specifiable states of the individual subject.⁶ In denying that psychology is individualistic, Burge is claiming that the individuation of mental state types depends on the possessor's environment or social context, such that a person's mental states may vary with his environment or social context, while his physical, neurological, and functional states, i.e. his non-intentionally specifiable states, remain constant. If Burge is right that contextual features play a role in determining the content of mental states, and that psychology type-individuates mental states by their content, then psychology seemingly violates the formality condition, since the formality condition requires that the content of mental states supervenes on formally specifiable properties of the individual subject.

What is Burge's evidence for the claim that psychology is non-individualistic? In particular, do the Putnam and Burge thought experiments establish that individualism in psychology is false? Even if the thought

experiments strongly suggest that attributions of propositional attitudes in everyday discourse are not individualistic (a question to be examined below), one might hold that theoretical psychology is or should be individualistic, relying on theoretical constructs different from the notion of content employed by folk psychology. Burge, however, argues that the methods and presuppositions of psychology, as it is currently practiced, are non-individualistic, and are likely to continue to be so. Burge claims that in describing and explaining the activities of cognitive agents, psychology makes use of interpreted that-clauses and other intentional constructions which are subject to the thought experiments. This use, he claims, is not merely heuristic or instrumental; furthermore, there is no well-articulated individualistic language which could replace the non-individualistic language that psychology is currently committed to.

Burge's arguments against individualism in psychology are based on claims about how psychology is actually practiced. I shall discuss Burge's arguments at length below. First, however, I would like to consider an argument that purports to establish that individuation in psychology must be individualistic if psychology is to count as a science.

1. An A Priori Argument for Individualism

Jerry Fodor, in Psychosemantics,⁷ gives an a priori argument for the claim that psychology should be individualistic. According to Fodor, "individuation in science is always individualistic."⁸ If psychology is to be scientific, he concludes, it should taxonomize mental states individualistically. Furthermore, he claims, "if mind/brain supervenience goes, the intelligibility of mental causation goes with it."⁹ I shall argue that Fodor's claims about individuation in science are false, at least outside the basic sciences of physics and chemistry. A non-individualistic psychology need not be "unscientific", nor must it make a mystery of mental causation. In what follows I shall not argue that psychology does or should taxonomize Burge's way; I am concerned merely to show that Fodor's a priori argument does not establish that psychology should be individualistic.

A few remarks about terminology: I shall use the term "individualism" as Burge uses it - to refer to a thesis about individuation in psychology (in particular, the thesis that mental states supervene on non-intentionally specified states of the individual subject). Fodor uses the term somewhat differently. He defines methodological individualism as "the doctrine that psychological states

are individuated with respect to their causal powers."¹⁰ This thesis, I shall argue, is compatible with anti-individualism of the Burgean sort (i.e. non-supervenience). To add to the terminological confusion, Fodor sometimes talks of non-psychological theories being individualistic (as in the claim that individuation in science is always individualistic) although he does not explicitly define a notion of individualism that applies to all of science. Fodor's claim that individuation in science is always individualistic is most plausibly interpreted as the claim that scientific theories individuate states and entities by reference to their local micro-structure. Anti-individualism (of the Burgean sort) is incompatible with this thesis.

Fodor begins with an analogy. He defines two predicates, "is an H-particle at t" and "is a T-particle at t" such that a particle is either an H-particle or a T-particle at a particular time depending upon whether a particular dime of Fodor's is heads up or tails up at that time. By simply turning over his dime, Fodor can change every particle in the universe from an H-particle to a T-particle, and back again. He claims that a particle physics that counted such relational predicates among its explanatory resources would be "mad" because "particle physics, like every other branch of science, is in the business of causal explanation; and whether something is

an H- (T-) particle is irrelevant to its causal powers." 11

Fodor's argument depends on the claim that scientific theories individuate entities by their causal powers.^{12,13} Two things have the same causal powers if and only if they have the same causal consequences across nomologically possible contexts.¹⁴ The problem with H- and T- particle theory, according to Fodor, is that it type-distinguishes entities which have the same causal powers. A genuinely scientific theory, Fodor claims, would type-identify two objects which differ only in that one is an H-particle and the other is a T-particle.

The mental states of the two protagonists in the thought experiments are alleged to be analogous to two particles which differ only in that one is an H-particle and the other is a T-particle. Fodor claims that the mental states of the two twins when they utter the words "I desire a glass of water" have the same causal powers. Insofar as psychology is interested in causal explanation, he concludes, it should type-identify their mental states. And if, as Fodor is willing to concede to Burge, common-sense psychology type-distinguishes the two states,¹⁵ then the individuation criteria employed by scientific psychology and by its folk counterpart are just different.

Fodor considers two objections to the above argument. According to the first objection, the mental states of the

twins have different causal powers: mine have causal powers with respect to H₂O and my twin's have causal powers with respect to XYZ. Fodor responds that, nonetheless, the causal powers of the two states are the same; causal powers are to be compared across contexts, not within contexts:

What is relevant to the identity of causal powers is the following pair of counterfactuals: (a) If his utterance (thought) had occurred in my context, it would have had the effects that my utterance (thought) did have; and (b) if my utterance (thought) had occurred in his context, it would have had the effects that his utterance (thought) did have. For our utterances (thoughts) to have the same causal powers, both of these counterfactuals have to be true.

16

Fodor claims that both of the counterfactuals are true, because if I say "bring water" on twin-earth my utterance produces XYZ, and if my twin says "bring water" here her utterance produces H₂O. So, the two mental states seem to have the same effects when evaluated relative to the same contexts. It doesn't follow, however, that the two mental states have the same causal powers. Since, according to Fodor, two states have the same causal powers if and only if they have the same causal consequences in all nomologically possible worlds, the two counterfactual conditionals in the passage quoted above provide only a necessary condition for the identity

of causal powers of the twins' mental states, not a sufficient condition. If there exists a nomologically possible world where the utterance of the form of words "bring water" by an earthian and by a twin-earthian have different effects, then the two utterances (thoughts) have different causal powers. The following would seem to be such a context: suppose most twin-earthlings know that XYZ does not occur on earth. Although they cannot refer to water -- never having been in causal contact with it -- they do know that the stuff earthlings call 'water' is chemically different from their own 'water'. When a visiting monolingual earthling says "bring water" the twin-earthlings do nothing -- they know that the earthling is asking for something non-existent on twin-earth. A request for 'water' by a twin-earthling in the same context, of course, produces XYZ. The two utterances (thoughts), therefore, have different effects in this context. It would seem, then, that a theory which individuates states by their causal powers should count the two states as type-distinct.

The first objection to Fodor's argument that psychology should individuate mental states individualistically claims that the subjects' mental states have different causal powers by virtue of having different referents. A second objection claims that the subjects' mental states have different causal powers by

virtue of producing different behaviours, where behaviour is construed intentionally. For example, my desiring something to drink causes me to ask for a glass of water. My twin never asks for water -- her thirst produces a request for a different liquid. Our bodily movements in producing these behaviours are the same, of course; but construals of behaviour as bodily movements serve little purpose in psychology; the relevant construals of behaviour for psychology are intentional.

Fodor has two replies to the second objection. His first strategy is to argue that if behaviour is to be individuated such that the twins' behaviour counts as type-distinct, then not only should we conclude that the twins' mental states are different -- by parity of argument we should also conclude that their brain states are different:

My twin is in a brain state that eventuates in his uttering the form of words "Bring water". I am in a brain state that eventuates in my uttering the form of words "Bring water"; If our uttering these forms of words counts as our behaving differently, then it looks as though our brain states differ in their behavioural consequences, hence in their causal powers, hence in the state types of which they are tokens. 17

Fodor's argument is intended as a reductio ad absurdum of the claim that the twins' behaviour is type-distinct. It can be reconstructed as follows:

- (1) Suppose we count the behaviour of the twins as type-distinct.
- (2) Behaviour is a causal consequence of brain states.
- (3) So, the causal powers of the brain states of the twins are different. (from 1 and 2)
- (4) Scientific theories individuate states by their causal powers. (general principle)
- (5) So, the brain states of the twins are type-distinct. (from 3 and 4, together with the fact that neurology, the theory which individuates brain states, is a scientific theory)
- (6) But it is absurd to think that the brain states of the twins are type-distinct simply because they live in worlds with different liquids.¹⁸
- (7) Hence, we should not count the behaviour of the twins as type-distinct.

There are a number of independent problems with Fodor's argument. I shall argue below that the general principle that scientific theories individuate entities by their causal powers (line 4) needs to be qualified. First, though, a problem with the inference from (1) and (2) to (3) can be brought out by a simple example. Consider two physically identical boxes (B_1 and B_2), each rigged up with an amplifier and speaker. Machines of this type emit sounds under various specifiable conditions. Suppose that B_1 in the relevant conditions emits sounds at

the pitch middle C. B_2 , however, is situated in an environment with a denser atmosphere and in the relevant conditions it emits sounds pitched an octave lower. Now consider the following argument:

(1a) Suppose we count the sounds emitted by the two boxes as type-distinct.

(2a) The emitted sounds are causal consequences of the physical states of the boxes.

(3a) So, the causal powers of the physical states of the two boxes are different.

The argument could be continued, in the fashion of Fodor's argument, to the conclusion that the physical states of the boxes should be type-distinguished, but since it would be absurd to do so simply because their environments are different, we should not type-distinguish the sounds emitted by the boxes.

There is clearly something wrong with the above argument. The two boxes are by hypothesis physically identical; they therefore have the same causal powers. But does it follow that the boxes' "behaviour" (the sounds emitted) must be type-identified? Surely not. A difference in their environments, not a difference in their causal powers, accounts for the fact that the boxes produce different sounds. But whatever the reason, we do type-distinguish the sounds; the type-individuation of sounds is sensitive to factors other than the causal

powers of the thing producing the sounds.

If the principle at line 4 is correct then our taxonomic practices with respect to sounds are not scientific, but the argument is intended to establish something quite different - that type-distinguishing the sounds is inconsistent with type-identifying the physical-states (or causal powers) of the boxes that produce the sounds. No such inconsistency has been established, nor is the claim independently plausible, given that both individuating practices are well-entrenched.

The same can be said of the behaviour of the twins. The issue in the present context is not whether a practice which type-distinguished the twins' behaviour is scientific, it is whether it is consistent with type-identifying their brain states. Fodor has offered no reason to think that it isn't. Anti-individualists propose to type-individuate behaviour partly by reference to its context. In doing so they are denying that behaviour supervenes on the causal powers of brain states, that is, they are denying that the protagonists must be in different brain states. Whether or not such a position is "scientific" (an issue to be addressed below), it is certainly not incoherent.

The inference from (1) and (2) to (3) is not valid without an assumption that behavioural types supervene on brain types, which is the very question at issue. (1)

says that the twins' behaviour is type-distinct. The proper construal of (2) is that behavioural events (i.e. tokens) are causal consequences of brain states. To conclude that the causal powers of brain states (i.e. brain state types) are distinct is to assume that there can be no difference in behavioural type without a difference in brain state type, which is just to assume that behaviour (types) supervenes on brain states (types). This assumption, of course, is exactly what the anti-individualist denies.

To summarize: type-distinguishing the behaviour of the twins is claimed by Fodor to lead to the unacceptable conclusion that the twins are in different brain states. The conclusion of the reductio has been blocked: we need not type-distinguish the brain states of the twins on the assumption that their behaviour is different because the inference from type-distinct effects to type-distinct causal powers is not valid without the question-begging assumption that behavioural types supervene on brain types.

It might be noted that the failure of Fodor's argument should give no comfort to the anti-individualist. The original argument that the twins' mental states are type-distinct because their behaviour is type-distinct has the same form as both Fodor's argument and the "box" parody. All three arguments rely on the unsupported

assumption that type-distinct effects imply type-distinct causes. Even if the (intentionally construed) behaviour of the twins is type-distinct, it doesn't follow that their mental states have different causal powers, nor that their mental states should be type-distinguished. To argue that it does is to assume that behavioural types supervene on mental state types. This assumption needs independent argument, since an individualist might argue¹⁹ that the behaviours of the twins are type-distinct not because their mental states are type-distinct but because their environments are different.

Fodor has a second reply to the objection that since the behaviour of the twins, intentionally construed, is type-distinct, their mental states should be type-distinguished. He extends the analogy between non-individualistic psychology and H- and T-particle theory, pointing out that an H- and T-particle theorist might justify his individuating practices by arguing as follows: H- and T-particles do indeed have different causal powers -- H-particles enter into (cause) H-particle interactions and T-particles enter into (cause) T-particle interactions. According to Fodor, this move is analogous to the claim that the behaviour of the twins is type-distinct, hence their mental states are type-distinct. Both moves, he claims, are illegitimate.

Fodor is right about this. Both arguments commit the

fallacy discussed earlier -- assuming, without argument, that type-distinct effects imply type-distinct causes. From the claim that H- and T-interactions are type-distinct events it doesn't automatically follow that the particles involved in their causes are type-distinct. But let's consider what would be required to fix up the two arguments. In the twins case we need (1) an argument that the twins' behaviour is type-distinct, and (2) an argument that the type-distinct behaviours are best explained by type-distinct mental states (rather than by, say, aspects of their different environments). Analogously, the H- and T-particle theorist needs (1) an argument that "H-interactions" and "T-interactions" are indeed type-distinct events, and (2) an argument that these type-distinct events are best explained by reference to type-distinct particles which cause them. As I have been at pains to point out, both arguments lack (2).

What about (1)? Anti-individualists (notably Burge) have argued that folk psychology taxonomises behaviour in such a way that the behaviour of the twins would count as type-distinct. The argument for this claim will be discussed below; what is important in the present context is that such individuating principles are alleged to be part of an established practice. By contrast, the move by the H- and T-particle theorist to type-distinguish "H-interactions" and "T-interactions" is completely ad

hoc. In fact, the real problem with H- and T-particle theory, I suggest, is not that there is anything in principle wrong with the way it individuates entities (i.e. "relationally"), but that it seems to do no genuine explanatory work. The explananda of H- and T-particle theory "explanations" cannot be formulated independently of reference to the properties of being an H- or T-particle. Individuating entities as H- or T-particles seems silly because the theory sheds no light on any questions which can be motivated or formulated outside the theory, not because there is anything in principle wrong with individuating entities or states relationally.

I would now like to examine the principle at the heart of Fodor's argument that a scientific theory must taxonomize entities or states by their causal powers. Fodor's claims about individuating practices in science actually rest on two general principles:

Methodological point: Categorization in science is characteristically taxonomy by causal powers. Identity of causal powers is identity of causal consequences across nomologically possible worlds.

Metaphysical point: Causal powers supervene on local micro-structure. 20

The correct construal of the methodological point, I shall argue, is compatible with non-individualistic taxonomic principles. Fodor's metaphysical point, at least outside of the basic sciences of physics and

chemistry, is simply false.

The claim that science individuates states by their causal powers is false if taken to mean that a scientific theory individuates states by all their causal powers. Rather, scientific theories individuate states by their relevant causal powers, where relevant is specified by the theory in question. Other causal powers may simply be ignored because they are deemed irrelevant with respect to the explanatory concerns of the theory. For example, suppose that two subjects give similar responses in a series of psychological test situations, the only difference in their responses being a variation in their reaction times. Such a difference presumably indicates some neurological difference. It is not obvious, however, that a psychological theory must respect this difference in the causal power of the subject's brain states in type-individuating their mental states. The difference in reaction times may be construed as a measure of the complexity of underlying psychological (or computational) operations; or it may simply be deemed an irrelevant difference from the point of view of psychology. If the psychological theory type-identifies their mental states in spite of the neurological difference, then the subject's mental states will have the same causal powers, while their brain states have different causal powers. How is this possible, if a mental state just is a brain

state? The answer is that the causal powers of a state can only be specified relative to a particular way of individuating the state. Brain states and mental states, then, always have different causal powers. Causal powers are the properties in virtue of which events (states), so individuated, are subsumed under causal laws. The causal powers of mental states, qua mental states, are mental properties; they are just those properties of the event (state) in virtue of which it falls under psychological laws. The causal powers of brain states, qua brain states, are neurological properties; they are just those properties in virtue of which the event (state) falls under neurological laws. Since mental properties are not identical to neurological properties the causal powers of mental states and brain states are different.

The claim that science taxonomizes by reference to causal powers, then, needs to be qualified. Generally, only some of an event's (state's) causal powers are relevant to its type-individuation by a particular theory. Which of a brain state's causal powers are relevant to its type-individuation as a mental state depend in part on what the psychological theory is intended to explain.

The methodological point, by itself, doesn't specify what kinds of properties qualify as causal powers. Causal powers are just those properties in virtue of which a state (event) is subsumed under causal laws. The

stipulation that identity of causal powers is identity of causal consequences across nomologically possible contexts doesn't restrict the kinds of properties that can be causal properties. Since the causal powers of a state can only be specified relative to a particular way of individuating the state, the stipulation simply says that two states (events) are subsumed under the same causal laws by a theory just in case the states (events) have the same effects across nomologically possible contexts, when the effects in question are individuated in accordance with the theory. A psychological theory which type-distinguished the twins' mental states, thereby subsuming them under different causal laws, would also type-distinguish at least some of the effects of their mental states in some contexts, just in virtue of the fact that the states fall under different causal laws.

Fodor's metaphysical point, however, does put constraints on what properties can count as causal powers. The following is a plausible interpretation of the metaphysical point: where T1 is a theory which specifies the micro-structure of entities individuated by T2, T2 recognizes no difference in causal powers unless there is a corresponding difference in causal powers recognized by T1. In other words, the causal powers postulated by macro-level theory supervene on the causal powers postulated by the micro-theory. If Fodor's "metaphysical

point" is correct, it would seem to rule out a difference in the causal powers of subjects' mental states without a difference in the causal powers of their brain states, that is, it would seem to rule out type-individuating mental states as the anti-individualists suggest.

The conjunction of Fodor's two claims entails that individuation in science always supervenes on local micro-structure. Is this claim supported by scientific practice? In fact, it is not hard to find counterexamples to Fodor's sweeping claim. Fodor even produces some himself when he says

...the sciences are forever cross-cutting one another's taxonomies. Chemistry doesn't care about the distinction between streams and lakes, but geology does. Physics doesn't care about the distinction between bankers and butchers, but sociology does....things in Nature overlap in their causal powers to various degrees and in various respects; the sciences play these overlaps, each in its own way.

21

By parity of argument, neurology doesn't care about the distinction between the twins' mental states, although psychology might. Notice that what is common to the above examples is that the principle that individuation supervenes on local micro-structure is violated: it is claimed that taxonomy in geology and sociology do not supervene on taxonomy in chemistry and physics respectively. To say that chemistry doesn't care about

the distinction between streams and lakes but that geology does is to say that geology recognises distinctions that are not grounded in local micro-structure. In fact, many scientific theories employ individuating principles which take account of features of the history or environment of the states and objects they describe.²² Fodor does not try to argue that geological theories are any less scientific for the fact that they individuate entities non-locally (i.e. non-individualistically). Nonetheless, he does argue that a scientific psychology, to be scientific, must not violate individualism. But surely what's good enough for geology should be good enough for psychology.

Since causal powers are the properties of an event or state in virtue of which it is subsumed under a causal law, to say that scientific theories taxonomize by reference to causal powers is simply to say that scientific theories postulate causal laws. Fodor's worry, and the intuition underlying his "metaphysical point", seems to be that if a theory does not taxonomize by reference to local micro-structure, as non-individualistic psychology does not, its "laws" are not causal laws. Indeed, he says "if mind/brain supervenience goes, the intelligibility of mental causation goes with it."²³ According to Fodor, for the mental states of twins to have different causal powers there would have to be a mechanism

connecting (the causal powers of) each subject's mental state with relevant features of the environment, without affecting the subject's physiology (since the two subjects are, by hypothesis, physiologically identical). But, Fodor claims, "there is no such mechanism; you can't affect the causal powers of a person's mental states without affecting his physiology."²⁴ To avoid postulating "crazy causal mechanisms" (or action at a distance), he concludes, we must say that the twins' mental states have the same causal powers, and any scientific psychology should type-identify them.

The idea seems to be that if the twins' mental states are type-distinct (have different causal powers) in virtue of the fact that they live in worlds with different liquids, then the liquids (H₂O and XYZ respectively) must cause their different mental states. But the only way the environment can affect their minds is by affecting their bodies; since the twins are, by hypothesis, physically identical, to claim that their mental states are type-distinct is to invoke "action at a distance".

Burge²⁵ has pointed out that Fodor's reasoning confuses causation with individuation. The proper construal of the anti-individualist claim is not that differences in the environments cause a difference in the twins' mental states, but rather that environmental factors are relevant to the type-individuation of their

mental states. Our theories and the taxonomies employed by them are not caused by the world in any literal sense; questions of explanatory comprehensiveness, simplicity, etc. always play a role in determining a taxonomic scheme. In response to the charge of conflating causation and individuation, Fodor claims that a theorist cannot both individuate non-locally (i.e. take into account features of context) and individuate by causal powers. Fodor's response rests on the cogency of his "metaphysical point", that causal powers supervene on local micro-structure. But we have seen that scientific theories sometimes individuate non-locally. There is no reason to think that the laws of biology and geology are not causal: they certainly purport to be, and they do not assume any crazy causal mechanisms or action at a distance. The "metaphysical point" is simply unsupported by scientific practice, at least outside the basic sciences of physics and chemistry. Without the metaphysical point, however, Fodor's case against anti-individualist psychology collapses: the principle that science individuates by reference to causal powers is not itself incompatible with anti-individualism.

In summary, Fodor's argument that a scientific psychology should be individualistic rests on dubious principles not well-supported by scientific practice. The issue between Fodor and Burge cannot, I believe, be

settled by appeal to general metaphysical or methodological principles.

2. Does Psychological Practice Support Individualism?

I shall take it as established that no a priori argument can show that taxonomizing entities non-individualistically is "unscientific" or otherwise methodologically suspect. In the rest of this chapter I shall consider whether there are any compelling reasons to think that psychology, in particular, either should or should not be individualistic in its taxonomic practices. I shall concentrate on Tyler Burge's arguments in "Individualism and Psychology" for the claim that much of psychology is non-individualistic.

Before getting to Burge's arguments I would like to consider an example of Ned Block's which Fodor claims shows that psychologists do in fact individuate mental states individualistically:

...imagine a psychologist (call her Psyche...) who is studying the etiology of food preferences, and who happens to have both Oscar and Oscar² in her subject population. [26] Now, on the intuitions that Burge invites us to share, Oscar and Oscar² have different food preferences; what Oscar prefers to gruel is brisket, but what Oscar² prefers to gruel is brisket²....If [Psyche] discounts Oscar and Oscar², she'll be able to

say -- as it might be -- that there are two determinants of food preference: 27.3 percent of the variance is genetic and the remaining 72.7 percent is the result of early training. If, however, she counts Oscar and Oscar2 in, and if she counts their food preferences the way Burge wants her to, then she has to say that there are three sources of variance: genetic endowment, early training, and linguistic affiliation. But surely it's mad to say that linguistic affiliation is per se a determinant of food preference; how could it be?

27

Fodor goes on to suggest that Psyche ought to preserve her hypothesis that there are only two determinants of food preference by counting the two Oscars as having the same preferences. Fodor's argument again depends upon conflating causation and individuation. He says "if there is such a difference [in the Oscars' preferences], then there must be some mechanism which can connect the causal powers of Oscar's mental states with the character of his linguistic affiliation without affecting his physiological constitution."²⁸ But a psychologist who type-distinguishes the Oscars' preferences on the grounds that they are members of different linguistic communities is not thereby postulating linguistic affiliation as a third cause of their preferences, so there is no need for a mechanism to connect the linguistic community with the subject's mental states. Rather she is taking linguistic affiliation to be a factor relevant for the individuation of preferences.

To count linguistic affiliation as relevant for individuating preferences is compatible with holding that there are only two causes of current preferences - genetics and early training.

If a psychologist relies upon a subject's verbal reports as evidence of his preferences, then she must understand the commitments of the subject's language in specifying what his preferences are. Suppose one of Psyche's subjects speaks an old English dialect, and understands "meat" to include nuts. Clearly it would be incorrect for her to construe this subject's claim that he dislikes "meat" as expressing the same preference as expressed by a homophonic utterance by a modern English speaker. (To type-identify their preferences is to miss an important difference between the two: one shuns nuts, the other doesn't. The difference in their food-directed behaviour can only be explained by a difference in their preferences, which in turn might best be explained by some difference in genetics or early training. Since the subject's language obviously determines how he describes his preferences, a theorist who relies on verbal reports as evidence of preferences must understand the subject's language to know what preferences are being expressed. It is a further question, though, how the theorist chooses to describe the subject's preferences. She can preserve distinctions in the subject's language in her own

theoretical vocabulary if she chooses, or she can ignore them. What she will do will depend, in part, on her explanatory goals, e.g. is she trying to explain food-directed behaviour? The best individuating scheme will be one which allows her to capture interesting generalizations, given her explanatory goals.

Fodor's argument has the appearance of resting on observed features of psychological practice. It doesn't, and not simply because psychologists do not encounter subjects who are physically and functionally identical. It rests on Fodor's worries about "action at a distance" which I have argued are unfounded. Psychologists do hypothesize that preferences are caused by genetics and early training, but this hypothesis is consistent with non-individualistic type-individuation of preferences.

* * *

Burge has two arguments from psychological practice which purport to show that individuation in psychology is sometimes non-individualistic. His first line of argument depends upon the fact that psychology attributes intentional mental states to agents using interpreted that-clauses which are subject to the thought experiments:

I am assuming that [psychology] seeks to refine, deepen, generalize, and systematize

some of the statements of informed common sense about people's mental activity....In describing and, at least partly, in explaining these activities and abilities, psychology makes use of interpreted that-clauses and other intensional constructions -- or what we might loosely call "intentional content". I have seen no sound reason to believe that this use is merely heuristic, instrumentalistic, or second class in any other sense. 29

...the generalizations with counterfactual force that appear in psychological theories, given their standard interpretations, are not all individualistic. For ordinary understanding of the truth conditions, or individuation conditions, of the relevant attributions suffices to verify the thought experiments. 30

...if psychology did take the individualistic route suggested by [functionalist] approaches ...then its power to illumine the everyday phenomena alluded to in everyday discourse would be correspondingly limited. 31

Burge's argument depends upon the fact that that-clauses are our primary means of ascribing intentional mental states, in both psychological theory and everyday discourse. The thought experiments are then alleged to show that content, as individuated by psychological explanation, is non-individualistic.

Burge's argument is inconclusive. Even granting (1) that the thought experiments do show that environmental and social factors play a role in (oblique or de dicto) attitude ascriptions, and (2) that psychology makes use of interpreted that-clauses in ascribing propositional attitudes, it does not follow that psychology individuates

mental states partly by reference to environmental or social factors. Brian Loar³² has argued that the content of mental states, as individuated by commonsense psychological explanation (= psychological content), is not, in general, to be identified with what is picked out by oblique attitude ascriptions (= social content). Loar uses a number of examples to show that sameness of de dicto or oblique attitude ascription is neither necessary nor sufficient for sameness of psychological content.

Loar's first set of examples is designed to show that sameness of oblique attitude ascription is not sufficient for sameness of psychological content. He asks us to consider an English speaker, Paul, who believes that he has arthritis in his ankles. Paul visits Paris and hears of a disease called "arthrite". Not realizing that "arthritis" and "arthrite" are intertranslatable, and being a bit of a hypochondriac, Paul comes to believe that he has two problems with his ankles. The ailment he calls "arthrite" he mistakenly believes to have infected his thigh. Loar concludes that Paul has two beliefs concerning his ankles, each with distinct psychological contents, despite the fact that only one oblique belief ascription is true of him (viz. he believes that he has arthritis in his ankles). Loar argues as follows:

Had he been less inclined to hypochondria, his English belief might have been instead that he

does not have arthritis in his ankles. Now that belief would clearly have been psychologically consistent with his French belief that he has arthritis in his ankles, but not with his actual English belief. So the latter two have to be distinct in content....

33

Loar uses a second type of example to argue the converse -- that sameness of oblique attitude ascription is not necessary for sameness of psychological content. We are asked to suppose that we have found a diary which we know is by either an earthling or a twin-earthling, although we do not know which. One entry says: "No swimming today; we think the water is too rough". According to Loar, the reported psychological explanation loses nothing from the fact that we do not know whether the diarist is talking about water or some other liquid. We understand the psychological explanation, Loar claims, because we know how the diarist personally conceives things. Whether he is an earthling or a twin-earthling, the psychological content of his mental state is the same.

The feature of beliefs that determines their type-individuation by commonsense psychology, according to Loar, is their pattern of actual and potential interaction with other beliefs, with stimuli, and with behaviour, that is, their conceptual (functional) role. It is only as individuated by their conceptual roles, Loar argues, that mental states are subsumed under the generalizations that

constitute commonsense psychological theory:

Commonsense psychological explanation involves various elementary structures in the relations among beliefs, wants and so on. There are motivational structures: x's belief that p, x's belief that doing-A-brings-about-g-if-p-is-true, and x's desire that g often conjointly explain x's doing A. There are inferential structures: x's believing that p only if q and x's belief that p may explain x's belief that q.... These structures apply to beliefs and desires only as they are appropriately individuated. The simple cases that I have been discussing can be spun out in obvious ways to show that the appropriate individuation conditions are not captured by oblique readings of ordinary belief ascriptions.

34

The grounds for attributing to Paul two type-distinct beliefs about his ankles is that his French belief and his English belief have different interactive potential. To use Fodor's terminology, they have different causal powers. Paul's French belief that he has arthritis in his ankles does not interact in the appropriate way with his English belief that if he has arthritis in his ankles he should take aspirin, although his English belief that he has arthritis in his ankles does. The explanatory schemas of commonsense psychology only subsume Paul's mental states as they are individuated by their conceptual roles.

Consider once again the diary entry "No swimming today; the water is too rough." As commonsense psychologists (which we all are, at least in our

non-philosophical moments) we would explain the diarist's behaviour as follows: because the diarist believes that the water is rough and believes also that if the water is rough then one should not swim, the diarist believes that one should not swim. If our explanation is challenged on the grounds that we don't know whether the diarist has any beliefs about water at all, then, if Loar is right, we would likely produce a paraphrase of the original explanation that makes no commitment about the kind of liquid the diarist is talking about (e.g. "the diarist believes that the local sample of the liquid that fills the oceans and lakes is rough..."); that is, some single explanation that would allow us to subsume the beliefs of both earthlings and twin-earthling, rather than two explanations, one adverting to water and the other to XYZ. On Loar's view, our original commonsense psychological explanation attributes to the diarist only a certain way of conceiving of things; in particular, it attributes mental states that are inter-related (and related to stimuli and behaviour) in specified ways, but involves no commitment about the environment or social context in which the beliefs are embedded.

According to Loar, conceptual role determines a kind of "narrow" content which plays a prominent role in commonsense psychological explanation. In response to the objection that there is no way to specify narrow contents,

—since that-clauses are shot through with environmental and social presuppositions, Loar argues that we are able to individuate mental states for the purpose of psychological explanation without a means of specifying psychological content directly. We get at them in context through various devices, e.g. we report a person's words along with other utterances or facts that help us to interpret his words, as in the description of the examples. If twin cases (i.e. individuals identical, or more plausibly, similar in physical make-up or functional organization) were to become common then psychology would probably develop precise, more convenient ways of picking out mental states (or their conceptual roles), but for now, outside the philosopher's laboratory, that-clauses serve well enough.³⁵

It is not my concern to argue that all psychological explanations subsume mental states in virtue of their conceptual roles, although I think Loar has assembled a persuasive case that some do. The point I want to make is a more general one. Burge has incorrectly assumed that just because the language we use to refer to mental states in psychological explanation involves social and environmental presuppositions, that what is referred to is itself individuated by or specific to the environmental or social context. Properties of belief ascriptions (that-clauses) should not be attributed without further

argument to the mental states that they serve to pick out in psychological theory, especially since it is a widely acknowledged fact that belief ascriptions serve purposes in everyday discourse besides predicting and explaining the behaviour and cognitive capacities of the subjects to whom they are ascribed. Loar's examples show that, sometimes psychological theory individuates mental states individualistically even though the language it uses to refer to these states is not itself individualistic. The thought experiments themselves, then, and the fact that psychology uses that-clauses in ascribing intentional content, do not establish that psychology is non-individualistic in its taxonomic practices.

* * *

Burge's second line of argument appeals to David Marr's theory of early vision in support of the claim that individuating practices in psychology involve presuppositions about the specific nature of the subject's environment. Perceptual states are alleged to be individuated by reference to the distal stimulus which gives rise to the image that forms the input to the visual system.

I shall briefly sketch Marr's theory³⁶ before turning to Burge's argument. The theory of vision, as conceived

by Marr, is deployed at three different levels of description: (1) the computational theory, which specifies the function computed by the visual system, (2) the representation and algorithm, which specifies the representations which serve as inputs and outputs to the computational processes which implement the function, as well as the algorithms which describe these processes, and (3) the hardware implementation, which characterizes the physical realization of the representations and algorithms. Marr stresses the importance of the computational theory, arguing that the nature of the computational processes underlying vision depend more upon the computational problems to be solved than upon the hardware in which the solutions are realized.

The goal of the visual system is to derive a representation of three-dimensional shape from information contained in two-dimensional images. Marr's theory divides this task into four distinct representational stages; corresponding to each is a type of representation, tokens of which serve as the inputs and/or outputs to the computational processes that effect the derivation of shape information from the image.

The image serves as the input to the visual system and represents the intensity of illumination at the retinal surface. Its primitives are the intensity values at each point in the image.

84

The primal sketch makes explicit certain information in the image, in particular, information about intensity changes which will be used by later processes. The primitives of the primal sketch are such things as zero-crossings, blobs, terminations, edge segments, virtual lines, groups, boundaries, and curvilinear organization, all abstract properties of the image.

The 2.5-D sketch makes explicit the orientation and depth of visible surfaces of the distal scene, and contours of discontinuities in these surfaces, using a viewer-centered coordinate frame. The 2.5-D sketch is the output of early visual processes. Its primitives include local surface orientation, distance from viewer, discontinuities in depth, and discontinuities in surface orientation.

The 2.5-D sketch serves as the input to subsequent processes that construct the 3-D model representation, which describes shapes and their spatial organization in an object-centered coordinate frame, employing a hierarchical representation that includes volumetric and surface primitives.

The computational processes to which tokens of these representations serve as inputs and/or outputs, up to the construction of the 2.5-D sketch, are data-driven (bottom-up) processes which rely only on information contained in the images. The visual system makes use of

85

certain very general assumptions about the environment that reflect physical constraints on the pairing of retinal images with distal shapes; these assumptions provide the additional information necessary to recover shape from the information contained in the images. An example of such an assumption is Shimon Ullman's rigidity assumption, which says that "any set of elements undergoing a two-dimensional transformation that has a unique interpretation as a rigid body moving in space is caused by such a body in motion and hence should be interpreted as such."³⁷ Ullman has proved that three distinct orthographic views of four non-coplanar points in a rigid body are sufficient to determine its three-dimensional structure (the structure from motion theorem). In a world like ours where most things are rigid, a process that incorporated the rigidity assumption would generally be successful in recovering the three-dimensional structure of distal objects from three such views. Other processes incorporate similar general assumptions to recover information about shape from properties of the image.

I turn now to Burge's argument that Marr's theory of vision is non-individualistic. He argues as follows:³⁸

- (1) The theory is intentional.
- (2) The intentional primitives of the theory and the information they carry are individuated by reference to

contingently existing physical items or conditions by which they are normally caused and to which they normally apply.

(3) So if these physical conditions and, possibly, attendant physical laws were regularly different, the information conveyed to the subject and the intentional content of his or her visual representations would be different.

(4) We can conceive of relevantly different physical conditions and perhaps relevantly different (say, optical) laws regularly causing the same non-intentional, individualistically individuated physical regularities in the subject's eyes and nervous system.

(5) In such a case (by (3)) the subject's visual representations would carry different information and have different representational content though the person's whole non-intentional physical history (at least up to a certain time) might remain the same.

(6) The theory individuates some perceptual states in terms of their informational or intentional content.

(7) Therefore, individualism is false for the theory of vision.

Step one, Burge claims, is "evident" from the fact that the "top levels" of the theory are formulated in intentional terms.

Step two is supported by a number of examples from

the theory, one of which describes how general physical assumptions constrain the choice of the representational primitives out of which the primal sketch is constructed. According to Burge, "the intentional content of representations of edges or generalized cones is individuated in terms of specific reference to those very contingently instantiated physical properties, on the assumption that those properties normally give rise to veridical representations of them".³⁹

Step three is supported by the above considerations regarding the individuation of intentional content, and by examples illustrating the explanatory method of the theory. (One example cites the role of physical constraints in visual processing; in particular, the constraints on matching that facilitate stereopsis.) According to Burge,

The methods of individuation and explanation are governed by the assumption that the subject has adapted to his or her environment sufficiently to obtain veridical information from it under normal conditions. If the properties and relations that normally caused visual impressions were regularly different from what they are, the individual would obtain different information and have visual experiences with different intentional content.

40

Step four asserts the existence of twin-earth type thought experiments. Step five follows from steps three and four. Step six is claimed to be "unproblematic".

I shall argue that Burge's argument is flawed at steps one, two, and six.

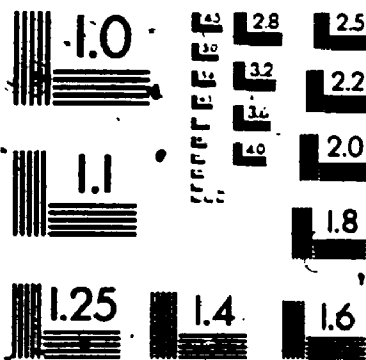
It is not clear what is being claimed in step one. Robert Matthews has argued persuasively⁴¹ that ascriptions of propositional attitudes play no essential role in Marr's theory. What Burge seems to mean by claiming that the theory is intentional is that the computational processes postulated by the theory are defined over representational tokens, and these representational states are semantically evaluable with respect to the distal scene. I shall return to this point below.

The argument breaks down at step two. Burge claims that the intentional primitives of the theory are individuated by reference to distal objects or features of the environment. In the first place, this is clearly not true of the inputs to the stereopsis and directional selectivity processes. These two processes take as inputs zero crossings⁴² from individual channels; zero crossings, Marr claims, are not "physically meaningful", in effect, they have no specifiable physical interpretation. Secondly, and more importantly, it is not obvious that the theory has any intentional primitives; what it does have are intentional specifications of primitives. Marr provides intentional specifications of the inputs and outputs of the function computed by the algorithms implementing the various processes (excepting stereopsis

2

of/de

2



MICRO

and ~~directional~~ selectivity), but the intentional specification of the computational theory (in the sense of Marr's top-most level) should not be confused with the computational theory itself.

The point can be sharpened by considering a single organism counterfactually situated in a number of distinct environments (i.e. the possibility considered by Burge in step four of the argument). There is no reason to think that the theorist would attribute to the subject's visual system a different function computed from one environment to the next. Since the algorithms implementing the computational processes do not change, there is no reason to assume that the function computed would change. What does change for each environment is the specification of the function computed.

But, if the same function is computed across different environments, then why should the specification of the function change from one environment to the next? Why does Marr not simply specify the function in some way that is neutral with respect to the particular environment in which the organism is situated? A plausible answer is that by specifying the representational tokens over which the computational processes of the theory is defined intentionally, in most cases as referring to properties of the distal scene, the theory makes these processes intelligible in the context in which the organism is

situated. A plausible adequacy condition on a theory of vision is that it contributes to our understanding of how the organism is successful in its environment to the extent that it is. Marr's theory meets this adequacy condition in part by specifying the function computed by the visual system by reference to general features of the organism's environment.

If the above is correct, then Marr's theory of vision is not intentional, although it does have intentional models. There is no reason to think that the primitives of the theory are intentional, although some of them have intentional specifications. Burge's argument, therefore, falls at steps one and two. It also fails at step six: the relevant perceptual states are specified in the theory's models by their intentional content in those models; the states themselves are not intentional.

Burge's argument can be seen as an instance of the same mistake that he made earlier in arguing for anti-individualism from the thought experiments. He falsely assumes that because the function computed by the visual system is specified in part by reference to general features of the organism's environment, that what function is computed is somehow determined by the environment.

To summarize what I have argued in this chapter: The

RTM proposes to individuate mental states by reference to their contents and respect the formality condition, the requirement that the contents of mental representations supervene on their formal properties. The Putnam/Burge thought experiments are alleged to show that contents do not supervene on formally specifiable properties of individual subjects, and consequently that psychological theories that individuate mental states by their contents will be non-individualistic, hence will violate the formality condition. I have argued that although Fodor's a priori argument fails to show that a "scientific" psychological theory must be individualistic, neither do arguments marshalled by Tyler Burge show that psychology is in fact non-individualistic. Of course, I have not argued, nor would I want to argue, that no existing psychological theories are non-individualistic. Some may well be; there are undoubtedly legitimate explanatory goals which would be well-served by non-individualistic theories, that is, by theories which ascribe "wide" contents to mental states.⁴³ I have simply tried to show that anti-individualism finds no support from either the thought experiments themselves or David Marr's theory of early vision.

Notes

1. Fodor, J.A. "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology," Behavioural and Brain Sciences, 3, 1980, pp.63-71.
2. See the discussion in chapter 1.
3. Putnam, Hilary. "The Meaning of Meaning," reprinted in Mind, Language, and Reality: Philosophical Papers Vol. 2, Cambridge: Cambridge University Press, 1975.
4. Burge, Tyler. "Individualism and the Mental," Midwest Studies, 4, 1979, pp.73-121. —
5. Burge, Tyler, "Individualism and Psychology," Philosophical Review, XCV,1, 1986, p.3.
6. Roughly, states of type A are supervenient on states of type B iff there can be no difference among A states without a corresponding difference among B states. Burge distinguishes the supervenience thesis from a stronger form of individualism which claims that mental states can be explicated by reference to physical or functional states of the individual. ("Individualism and Psychology", p.4).
7. Fodor, J.A. Psychosemantics: The Problem of Meaning in the Philosophy of Mind, Cambridge, Mass.: MIT Press, 1987.
8. Ibid., p.32.
9. Ibid., p.42.
10. Ibid. p.42.

11. Ibid., p.33.

12. Ibid., p.34.

13. In support of this claim Fodor says the following:

We want science to give causal explanations of such things (events, whatever) in nature as can be causally explained. Giving such explanations essentially involves projecting and confirming causal generalizations. And causal generalizations subsume the things they apply to in virtue of the causal properties of the things they apply to." (Ibid., 34)

Causal powers, then, are properties in virtue of which events are subsumed under causal laws.

14. Ibid., p.44.

15. On page 34, summarizing his argument, he says "... commonsense deploys a taxonomy that does distinguish between Oscar and Oscar²". (see fn. 26) I think this concession by Fodor is rash (see argument next section).

16. Ibid., p.35.

17. Ibid., p.37.

18. By hypothesis the twins are molecularly identical. In keeping with the rest of the twin-earth literature we are simply ignoring the fact that the brain is largely composed of water. The same point could be made using an example like gold -- the point being that the twins do not differ in their brain states simply because they are causally related to different substances in their

environments (or because they are members of different linguistic communities).

19. Fodor argues this position later in chapter 2.

20. Ibid., p.44.

21. Ibid., pp.44-45.

22. Burge, in "Individualism and Psychology," argues as follows:

A continent moves and is moved by local impacts from rocks, waves, and molecules. Yet we can conceive of holding constant the continent's peripheral impacts and chemically constituent events and objects, without holding identical the continent or certain of its macro-changes -- because the continent's spatial relations to other land masses affect the way we individuate it. Or take an example from biology. Let us accept the plausible principle that nothing causally affects breathing except as it causally affects local states of the lungs. It does not follow, and indeed is not true, that we individuate lungs and the various sub-events of respiration in such a way as to treat those objects and events as supervenient on the chemically described objects and events that compose them. If the same chemical process (same from the surfaces of the lungs inside, and back to the surfaces) were embedded in a different sort of body and had an entirely different function (say, digestive, immunological, or regulatory), we would not be dealing with the same biological states and events. (p.16)

23. Fodor, Psychosemantics, p.42.

24. Ibid., p.39.

25. Ibid., p.16.

26. In this variant on Burge's arthritis example, Oscar and Oscar₂ are two physically identical individuals who

differ only in the following respect: Oscar lives in a community which uses "brisket" to refer to the breast of any food animal, while Oscar² lives in a community which uses "brisket" to refer only to the breast of beef.

- 27. Fodor, Psychosemantics, p.40.
- 28. Ibid., p.40.
- 29. Burge, Tyler. "Individualism and Pyschology," p.8.
- 30. Ibid,, p.9.
- 31. Burge, Tyler. "Individualism and the Mental," p.109.
- 32. Loar, Brian. "Social Content and Psychological Content," unpublished ms.
- 33. Ibid:, p.10.
- 34. Ibid., pp.11-12.
- 35. What are that-clauses for?. Loar says:

That-clauses on their oblique readings are sensitive, either directly or indirectly via translation, to how beliefs would linguistically be expressed, and that is...only loosely related to psychological content. Now, as Burge's cases show, that-clauses capture how a belief would be expressed by exhibiting something that is equivalent in social content. (as we might say) to what the subject would utter, given his deference to the usage of his linguistic community... This enables that-clauses to capture certain extra-psychological relations of propositional attitudes to independent states of affairs, what we think of as their socially determined truth-conditions. The fundamental usefulness of this is that we may then describe people as conveyors of (more or less) determinate information, which remains constant even when the psychological contents

96

of their states vary. That-clauses enable us to impose a grid of socially regularized information on the vagaries of individual psychology. (pp.19-20)

36. Marr, David. Vision. New York: Freeman and Co., 1982.

37. Ullman, Shimon. The Interpretation of Visual Motion.
Cambridge, Mass: MIT Press, 1979.

38. Burge, Tyler. "Individualism and Psychology," p.34.

39. Ibid., p.35.

40. Ibid., p.33

41. Matthews, Robert. "The Alleged Evidence for
Representationalism," forthcoming.

42. A zero-crossing is a point where the value of a
function changes its sign. Zero-crossings correspond
to sudden intensity changes.

43. Preference psychology may be a case in point. (see
the discussion pp. 23-26) If Psyche's subject
population includes subjects from different linguistic
communities, then a non-individualistic taxonomy -- in
terms, say, of the foods that her subjects' behaviour
is directed towards -- would give Psyche more
comprehensive generalizations than would an
individualistic taxonomy, i.e. a taxonomy in terms of
the way the subjects themselves conceive of their
food-directed behaviour.

Chapter 4

Intermezzo: The Formality Condition Again

We assume that people do things because of their beliefs and desires, and commonsense psychology explains behaviour by reference to beliefs and desires; but at the same time, as good physicalists, we assume that mental processes are, at bottom, just causal sequences of states of the brain. Bringing these two ideas together is the major task of a theory of mind.

As we have seen, the RTM proposes to explain how propositional attitudes can be causally efficacious in the production of behaviour by respecting what has come to be known as the formality condition. As interpreted by the RTM, the formality condition requires that mental representations have their causal rôles in virtue of their syntax, or, put another way, that "the syntactic structure of mental representations reflects all relevant semantic distinctions."¹

Zenon Pylyshyn explains the basic idea as follows:

There is a fundamental difference between a description of a computer's operation cast in terms of its states (i.e. equivalence classes of physical descriptions) and one cast in terms of what it is about.... The fundamental difference is that the former refers to intrinsic properties of the device, while the latter refers to aspects of some entirely different domain, such as chess. The former can be viewed as a syntactic description,

while the latter is semantic, since it refers to the represented domain. 2

Because a computational process has no access to the actual represented domain itself (e.g. a computer has no way of distinguishing whether a symbol represents a number or letter or someone's name), it is mandatory, if the rules are to continue to be semantically interpretable (say as rules of arithmetic), that all relevant semantic distinctions be mirrored by syntactic distinctions -- i.e., by features intrinsic to the representation itself. Such features must in turn be reflected in functional differences in the operation of the device. That is what we mean when we say that a device represents something. Simply put, all and only syntactically encoded aspects of the represented domain can affect the way a process behaves. 3

The formalist view requires that we take the syntactic properties of representations quite literally. It is literally true of a computer that it contains, in some functionally discernible form... what could be referred to as a code or inscription of a symbolic expression, whose formal features mirror (in the sense of bearing a one-to-one correspondence with) semantic characteristics of some represented domain, and which causes the machine to behave in a certain way. 4

In the second quoted passage above, Pylyshyn points out that if a device is to be interpretable -- say, as an adding machine or a chess player -- then all relevant semantic distinctions must be preserved in the syntax of the device and syntactic distinctions must be preserved in its functional organization. This is achieved as follows: there is an interpretation function F_1 which maps features of the represented domain into features of the

representation itself, and a further interpretation function F_j which maps these syntactic features into functionally defined equivalence classes of physical features of the device. In other words, a syntactic description of a device is as much an interpretation of the device as is a semantic description. But then it is hard to understand why, in the first quoted passage, Pylyshyn says that the syntactic description refers to intrinsic properties of the device. Rather, it is the functional/physical description which refers to intrinsic properties of the device, the syntactic description being a product of interpretation. Pylyshyn's tendency to overlook the fact that the device has both its semantic and syntactic properties only under a particular interpretation of the device leads him to adopt a realist stance toward the device's syntactic properties which he does not extend to its semantic properties. In the third quoted passage he says that the syntactic description of the device is literally true. But surely the semantic and syntactic descriptions are on a par: each is true only under a particular interpretation of the device. And it is misleading to say that (third quoted passage) "all and only syntactically encoded aspects of the represented domain can affect the way a process behaves." What affects the way the process behaves are the physical states of the device; syntactic properties, like semantic

properties, affect the device's behaviour only under an interpretation mapping them into (functionally defined equivalence classes of) physical states of the device.

If Pylyshyn is willing to be realist about the syntactical description of the device, then he should be realist about its semantical description. Semantic and syntactic properties affect the device's behaviour only insofar as they are reflected in physical properties of the device. The formality condition, as interpreted by the RTM -- that mental representations have their causal roles in virtue of their syntax, or that semantics must supervene on syntax -- does not guarantee that states picked out by reference to their content will be causally efficacious; they won't be unless syntactic features also supervene on physical features of the device. The formality condition, then, appears to require that if semantic features are to be causally efficacious they must supervene on physical features of the device.

The RTM theorist might respond that the only way we know to make semantic features causally efficacious is to encode them in a syntax. The RTM, recall, construes thinking as mechanical theorem-proving, interpreting the sequence of events that causes behaviour as a computational derivation which has a description of the behaviour as its conclusion. RTM theorists often argue that this is the only way to understand cognitive

processes -- that it's "the only game in town". But this response should not be taken too seriously. Computational theory satisfies the formality condition by encoding semantic features in a syntax and building devices whose physical processes are sensitive to the encodings, but nature may do it differently, especially since we cannot assume that nature starts with an explicit description of some task domain (e.g. arithmetic, chess) and then tries to build a device that computes the function specified. The connectionist (or "parallel distributed processing"⁵) program assumes that intelligent machines can be built without syntactic encoding. It remains to be seen how fruitful connectionism will turn out to be; in any event, the only-game-in-town argument is undercut by the mere possibility of an alternative, even if it is too early to tell how well the new game will play.

The formality condition, I have argued, seems to require that content supervene not on syntax but on physical states of the device. I argued in chapter 3, however, that mental states individuated by their content can be causally efficacious in producing behaviour without such states supervening on brain states. What is required for the causal efficacy of the mental is that mental states, however individuated, can be realized as physical states of the device. The intelligibility of mental causation requires only the token-identity of mental

states and physical states. The formality condition imposes no constraints on psychological theory not already required by physical realizability.⁶ The formality condition allows that contents might be individuated widely (i.e. non-individualistically), and still be causally efficacious in producing behaviour, since mental states individuated by reference to wide contents are physically realizable as brain states.

Notes

1. Pylyshyn, Zenon W. "Computation and Cognition: Issues in the Foundation of Cognitive Science", The Behavioral and Brain Sciences, 3, 1980, p.114.
2. Ibid., p.113.
3. Ibid., pp.113-114.
4. Ibid., p.114.
5. See McClelland, J. and Rumelhart, D., eds. Parallel Distributed Processing: Explorations in the Microstructure of Cognition, 2 vols., Cambridge, Mass.: MIT Press, 1986.
6. Brian Smith argues that the formality condition is most plausibly construed as realizability in "Semantic Attribution and the Formality Condition", a paper delivered at the eighth annual meeting of the Society for Philosophy and Psychology, UWO, May, 1982. Ausonio Marras ("The Churchlands on Methodological Solipsism and Computational Psychology", Philosophy of Science, 52, 1985, p.305.) also holds that the formality condition imposes only a realizability constraint on psychological theories.

Chapter 5

Problems with the Representational Theory of Mind

The Representational Theory of Mind (RTM) holds that for any organism O and attitude A toward the proposition p , there is a computational relation R and a mental representation MP such that

MP means that p , and

O has A to p if and only if O bears R to MP .

The RTM further claims that mental processes are causal sequences of tokenings of mental representations.¹

The RTM has been explicitly offered as a vindication of the intentional realism inherent in commonsense psychology. Intentional realism, recall, is committed to the existence of internal states that (1) are semantically evaluable, (2) play a causal role in the production of behaviour, and (3) are truly described by the generalizations of commonsense psychology.

In this chapter I shall argue both that the RTM does not enjoy the support that is commonly claimed for it, and that its prospects for vindicating folk psychology are not as promising as has been claimed. In the course of criticizing the arguments alleged to support the RTM, I shall sketch the outline of an alternative account of propositional attitudes which, I shall argue, may stand a better chance of vindicating folk psychology.

1. Two Kinds of Counterexamples

Since the RTM is formulated as a biconditional it could be falsified in two ways -- either by cases where a subject has a propositional attitude but is not computationally related to an appropriate internal token, or by cases where a subject is computationally related to a contentful token in an internal language, but where there are no non-question-begging grounds for attributing a corresponding attitude.

Counterexamples of the latter sort would not necessarily undermine the RTM. If the theory was well-confirmed except for some apparent counterexamples of the second sort, the RTM theorist would probably be justified in claiming that a new kind of attitude had been discovered. As Fodor points out,² to vindicate commonsense psychology the RTM need not pick out a kind that is precisely coextensive with the propositional attitudes currently recognized by folk theory.

Counterexamples of the first sort would be more damaging. Dennett gives the following putative counterexample:

In a recent conversation with the designer of a chess-playing program I heard the following criticism of a rival program: "It thinks it should get its queen out early." This ascribes a propositional attitude to the program in a very useful and predictive way, for as the

designer went on to say, one can usually count on chasing that queen around the board. But for all the many levels of explicit representation to be found in that program, nowhere is anything roughly synonymous with "I should get my queen out early" explicitly tokened. The level of analysis to which the designer's remark belongs describes features of the program that are, in an entirely innocent way, emergent properties of the computational processes that have "engineering reality." I see no reason to believe that the relation between belief-talk and psychological-process talk will be any more direct. 3

Dennett's example raises the following problem for the RTM: there would seem to be compelling reasons for attributing to the device the belief that it should get its queen out early, despite the fact that no token with the appropriate content is explicitly represented in the device. Similar examples are not hard to find. A visual system that implements Shimon Ullman's structure-from-motion theorem might plausibly be said to assume that objects are rigid, even though the rigidity assumption is nowhere tokened in the system. In both cases the conditions that we normally take to be sufficient for ascribing a belief to a subject are satisfied. Not only does the attribution of the relevant attitude accurately predict the subject's behaviour, but we also seem to be as justified in taking the attitude to be causally implicated in the subject's behaviour as we ever are when we ascribe propositional attitudes.^{4,5} The relevant counterfactual conditionals seem to be satisfied (if the device didn't

believe it is a good idea to get its queen out early it wouldn't have made that sequence of moves...; if the organism didn't assume rigidity it wouldn't have been able to recover the shape of the object from such limited information...etc.). In neither case is there an explicit representation of the content of the attributed attitude. It looks like having the attitude isn't (always) a matter of being computationally related to a token mental representation. So the RTM is false.

Fodor's strategy in response to these apparent counterexamples is to appeal to a distinction between core and derivative cases of propositional attitude tokenings. What is required, according to Fodor, is a correspondence, in what the RTM takes to be the core cases, between the tokening of an attitude and the tokening of a corresponding mental representation. Of course, he points out, the RTM must provide independent, non-question-begging grounds for distinguishing core from derivative cases. It does so, he claims, as follows:

... mental processes are causal sequences of transformations of mental representations. It follows that tokenings of attitudes must correspond to tokenings of mental representations when they -- the attitude tokenings -- are episodes in mental processes. If the intentional objects of such causally efficacious attitude tokenings are not explicitly represented, then RTM is simply false...The motto is therefore No Intentional Causation without Explicit Representation. 6

Fodor's attempt to distinguish core and derivative cases of propositional attitude tokenings by appeal to the way the RTM explicates mental processes is simply question-begging. The RTM identifies mental processes with causal sequences of transformations of mental representations, so if the attitude tokenings in question do not figure in sequences of transformations of mental representations, then they are not episodes in mental processes. But what distinguishes all of the alleged counterexamples (of the Dennett sort) is precisely that the attitude tokenings in question do not figure in such sequences; this is why they are alleged to be counterexamples! To respond, as Fodor does, that these cases are merely derivative propositional attitude tokenings because they aren't episodes in mental processes (and they aren't episodes in mental processes because they do not figure in causal sequences of transformations of mental representations) is just to beg the question.

Fodor's response has other problems besides being question-begging. If the RTM is to vindicate folk psychology, then, as Fodor repeatedly stresses, folk psychological explanations of behaviour, when construed along the lines specified by the RTM, must turn out, for the most part, to be true. But this would seem to require that if folk psychology explains the behaviour of a system by ascribing (what the theory takes to be) causally

efficacious propositional attitudes to that system, then the RTM should take these propositional attitude tokenings to be core cases. The explanatory burden should be borne by computational states that are token-identical to the beliefs and desires that figure in folk psychological explanations. Otherwise the folk psychological explanations would not be true. This constraint is not satisfied in Fodor's construal of the Dennett example. According to Fodor, what must be explicitly represented in the device is not some content like "it's good to get the queen out early", but rather the actual and possible states of play of the game. The representations of the board must be explicit, he says, because the machine's computations are defined over them; transformations of token representations of board configurations constitute the machine's mental processes. But while the folk psychological explanation of the device's behaviour may attribute to it beliefs and desires about states of the board, these are not the propositional attitudes that bear the explanatory burden. The folk psychological explanation of the device's queen-directed behaviour would look something like the following: The device desires to win and believes that by getting its queen out early it increases its chances of winning. The device therefore desires to get its queen out early and acts accordingly. This generalization explains the device's behaviour in a

wide range of circumstances where the states of the board have nothing in common. Therefore, while the full explanation of the device's behaviour in a particular circumstance will appeal to beliefs and desires about states of the board, the generality of the explanation derives from the ascription of beliefs and desires about getting its queen out early. But, according to Fodor, these propositional attitudes are merely "derivative" -- their tokenings do not correspond to mental processes; yet they are the propositional attitudes that bear the explanatory burden in the folk psychological account. Fodor's response entails that the folk psychological generalization does not pick out the internal states of the device which are causally efficacious in producing its behaviour. Since the folk psychological generalization purports to be a causal explanation of the device's behaviour we should conclude that it is simply false. Folk psychology, it would seem, has the etiology wrong.

To summarize: Fodor appeals to a distinction between core and derivative cases of propositional attitudes tokenings in an attempt to insulate the RTM against Dennett-like counterexamples. But Fodor's response to Dennett's challenge is question-begging. His response has an additional cost: in disarming putative counterexamples in this way, it threatens the RTM's promise to vindicate folk psychological explanations of behaviour. For every

counterexample labelled a derivative attitude tokening, there corresponds a false folk psychological explanation. Fodor's strategy, therefore, threatens intentional realism: to the extent that we have reason to doubt that a device's mental processes include causal sequences of tokenings of mental representations with the content that p, we have reason to doubt whether our ascriptions of the belief that p or the desire that p to the device are true, no matter how well-confirmed such ascriptions might otherwise be. Insofar as we don't know how a device works, we don't know whether our propositional attitude ascriptions are true of it. Since we don't know how humans work, our confidence in the truth of folk psychological explanations of behaviour would seem to be misplaced.

2. Adequacy Conditions on a Theory of Propositional Attitudes

In the paper "Propositional Attitudes"⁷, Fodor cites five adequacy conditions on a theory of propositional attitudes:

(1) Propositional attitudes should be analysed as relations. Fodor adduces three considerations in favour of this condition. (i) "Believes" looks like a two-place

predicate, although he recognizes that appearances are sometimes deceiving (e.g. the idiom "lose your marbles" looks like a two-place predicate too); (ii) existential generalization applies to the syntactic objects of propositional attitude verbs (e.g. from "John believes it is raining" it follows "John believes something"); and (iii) the only alternative account, the "fusion" story, is seriously flawed.

The fusion story analyses the predicate "believes it is raining" as a one-place predicate with no internal structure. The sentence "John believes it is raining" is, on the fusion account, an atomic sentence, similar to "John is tall". The account has insurmountable problems besides being counter-intuitive: among other things, it doesn't get the individuation conditions of attitude ascriptions right, and it makes language learning a mystery.

(2) A theory of propositional attitudes should explain the parallelism between PA-verbs and verbs of saying. "...the things we can be said to believe (want, hope, regret, etc.) are the very things that we can be said to say (assert, state, etc.)."⁸

(3) A theory of propositional attitudes should account for their opacity. Existential generalization and substitution of co-referring expressions may fail inside belief contexts.

(4) The objects of propositional attitudes have logical form. According to Fodor, the form of the belief ascribed by a given belief sentence is determined by the logical form of the sentence that expresses the belief. For example, John's belief that Reagan and Bush are Republicans is a conjunctive belief; it has the same logical form as the sentence "Reagan and Bush are Republicans". Fodor says "A theory of propositional attitudes should legitimize the ascription of form to the objects of propositional attitudes. In particular, it should explain why the form of the belief is identical to the logical form of the correspondent of a sentence which (opaquely) ascribes that belief".⁹

(5) A theory of propositional attitudes should mesh with empirical accounts of mental processes. Fodor isn't clear on how strong he wants this requirement to be. He claims that a theory of propositional attitudes should explain in virtue of what facts propositional attitude ascriptions are true, but then he weakens the condition, claiming "It seems to me self-evident that no such theory could be acceptable unless it lent itself to explanations of the data -- gross and commonsensical or subtle and experimental -- about mental states and processes."¹⁰ The latter seems to require only that an account of propositional attitudes not be inconsistent with empirical accounts of mental processes, surely a reasonable

requirement.

A theory which construes propositional attitudes as relations between organisms and sentences of a natural language satisfies conditions (1)-(5), but is nonetheless untenable on other grounds. In the first place, it slices propositional attitudes too thin. For example, "John believes the dog bit Mary" and "John believes Mary was bitten by the dog" are different sentences but plausibly pick out the same belief. Furthermore, both English speakers and French speakers can surely share some beliefs. One solution to the thinness of slice problem is to take translation sets of sentences as the objects of belief, but this proposal encounters notorious problems of its own.¹¹ Secondly, construing beliefs as relations to sentences leaves animals and infra-linguistic humans out of the account. Thirdly, the theory raises questions about how a first language could be learned, since it would seem that language learning must itself presuppose the having of some propositional attitudes.

A theory which construes propositional attitudes as relations between organisms and sentences in an internal language, Fodor argues, satisfies the five initial adequacy conditions and avoids the above problems. It is assumed that the formulae of the internal language correspond one-to-one to the contents of propositional attitudes (e.g. "The dog bit Mary" and "Mary was bitten by

the dog" correspond to the same internal sentence). It is further assumed that the internal language is innate (so an organism can have propositional attitudes before it learns a natural language) and is as universal as human psychology; that is, "to the extent that an organism shares our mental processes, it also shares our system of internal representations."¹²

I shall argue that conditions (1) through (4) constrain an account of propositional attitudes only indirectly; they are more plausibly construed as adequacy conditions on an account of propositional attitude ascriptions. Consider Fodor's arguments for construing propositional attitudes as relational states: he first cites two striking properties of belief ascription -- that "believes" looks like a two-place predicate and that existential generalization applies to the syntactic objects of propositional attitude verbs -- and then criticises an alternative account, the 'fusion' story, that analyses propositional attitude verbs as one-place predicates with no internal structure. Fodor is clearly right that the fusion story is hopeless, but not because it is a false theory of propositional attitudes. It is not a theory of propositional attitudes at all, but rather a theory of propositional attitude ascription, and as such it is clearly false, for the reasons that Fodor adduces, (viz., it is counterintuitive, it doesn't get the

individuation conditions right, etc.). But from the fact that verbs of propositional attitude are exactly what they appear to be -- two-place relational predicates -- it does not follow that propositional attitudes themselves are relational states.

The point can be illustrated using an analogy of Paul Churchland's.¹³ He suggests that we think of the role of propositions in propositional attitude ascriptions as analogous to the role of numbers in physical magnitude ascriptions:

The idea that believing that p is a matter of standing in some appropriate relation to an abstract entity (the proposition that p) seems to me to have nothing more to recommend it than would the parallel suggestion that weighing 5 kg is at bottom a matter of standing in some suitable relation to an abstract entity (the number 5). For contents of this latter kind, at least, the relational construal is highly procrustean. Contexts like

x weighs 5 kg
 x moves at 5 m/s
 x radiates at 5 J/s

are more plausibly catalogued with contexts like

x weighs very little
 x moves quickly
 x radiates copiously.

In the latter three cases, what follows the main verb has a transparently adverbial function. The same adverbial function, I suggest, is being performed in the former cases as well. The only difference is that using singular terms for numbers in the adverbial position provides a more precise,

systematic, and useful way of modifying the main verb, especially when said position is open to quantification. 14

Churchland's immediate target is the view that to believe that p is to stand in some mysterious relation ("grasping") to an abstract entity (a proposition), but his analogy also challenges the view that propositional attitudes are relations between organisms and internal sentences. The point is that propositional attitudes need not be analysed as relations at all, despite the fact that propositional attitude ascriptions have the surface structure of two-place predicates. One should not, however, attach too much significance to Churchland's adverbial translations of relational contexts. The fact that 'believes that p ' and 'weighs x kg' can both be construed as predicate-forming functors turning singular terms into elements of complex predicates does not itself undermine the view that propositional attitudes are relational states. As Hartry Field¹⁵ has noted, insertion of a singular term in the second place of any two-place predicate results in a complex one-place predicate, so in a sense all relational sentences involve predicate-forming functors. The only obvious construal of the adverbial account that would constitute a genuine alternative to the relational account of propositional attitude ascriptions is the fusion view, which takes the syntactically complex

predicates to have no semantically significant constituent structure, and all parties agree that this view is hopeless.

What the Churchland analogy suggests is that propositional attitude ascriptions might be treated along the lines of physical magnitude ascriptions such as "has a mass of 10 kg," where we are not at all inclined to treat the designated physical magnitude as a relational property.¹⁶ In the latter case, the theory of measurement explains why the real numbers can be used to measure mass: a representation theorem can be proved such that if a system of objects has certain properties and relations, then there is a mapping of the system into the real numbers which preserves the structure of these properties and relations. Although assigning real numbers to the objects in the system is a convenient way to refer to their intrinsic mass relations, the intrinsic relations themselves do not depend on the existence of real numbers and are specifiable without reference to the real numbers.

According to Churchland, the role of propositions in propositional attitude ascriptions is analogous to the role of numbers in physical magnitude ascriptions. The causal relations holding among propositional attitudes have as a homomorph the logical relations holding among their contents. Assigning propositional contents to the internal states of the organism therefore allows us to

exploit the logical structure of propositional space in discussing the causal relations among the propositional attitudes themselves. I shall call this proposal the propositional measurement theory.¹⁷

Field, in the postscript to "Mental Representation", argues that the success of Churchland's propositional measurement proposal would actually support the internal representation hypothesis:

The "structure" that such a mapping would have to preserve would be the kind of structure important to propositions: viz. logical structure. So the system of entities inside the believer would have to have logical structure, and this I think means that the system of entities inside the believer can be viewed as a system of sentences -- an internal system of representation. In other words, I think it can be shown...that any system of entities that satisfies the condition of being related to the system of propositions via a structure-preserving map is a system of internal representation... 18

There are two possible interpretations of Field's remarks:

(1) the propositional measurement theory entails the existence of an internal system of representations because in fact only if propositional attitude states have sentential structure could they be successfully mapped to propositional space; or (2) the propositional measurement theory entails the existence of an internal system of representations because we should, as a matter of logic,

be willing to count as a system of representations any system of properties and relations that map into propositional space.

The first interpretation is perhaps the more plausible construal of Field's remarks, but it is unsupported. Field has no argument that postulating an internal system of representations is the only way that the mapping could be effected. In fact, it is somewhat implausible that propositional attitude states have sentential structure since empirical research shows that certain causal relations among propositional attitudes are not predicted by logical relations among their associated contents.¹⁹ This suggests that what is preserved under the mapping from internal states to propositional space is causal relations and that these relations are only represented in propositional space as logical relations. It suggests further that causal relations among propositional attitude states are preserved under the mapping to propositional space only under an idealization.

It is unlikely, then, that the postulation of an internal system of representations is the only way that the mapping could be achieved, or even that it is the most promising candidate. Development of the propositional measurement idea would involve the postulation of a system of properties and relations holding among internal states of the organism and between internal states and external

objects; where such relations are characterized independently of the propositional contents to which the internal states are mapped by the measurement theory. I shall call a theory which characterizes the properties and relations of internal states independently of reference to their propositional contents an intrinsic theory. An intrinsic theory need not postulate an internal system of representation, but might instead treat propositional attitudes as functional states of the organism which can, in principle at least, be specified independently of the propositional contents which index them.²⁰ Field has offered no reason to think that functional roles could not be successfully mapped into propositional space.

Given that Field has not ruled out an alternative to the internal representation hypothesis as a way of achieving the mapping between internal states and propositions, the second possible interpretation of his remarks -- that as a matter of logic the propositional measurement theory entails the internal representation hypothesis -- cannot be supported. It might possibly be claimed that any account which postulates a structure-preserving mapping from internal states to propositions thereby postulates an internal system of representation; in other words, the internal states themselves count as tokens in a language just in virtue of the fact that the causal relations holding among them have as a homomorph

the logical relations holding among propositions. But this construal of the internal representation hypothesis would empty it of its distinctive content. The RTM claims something much stronger -- that the internal states are themselves to be analysed as relations to representations. I take it that Field's remarks should not be interpreted as presupposing only the very weak construal of the internal representation hypothesis.²¹

To summarize the foregoing: Field seems to assume that the propositional measurement proposal would, in effect, treat propositional attitudes as relational states, in particular, as relations between organisms and internal sentences. This is one way that Churchland's proposal could be developed, but an intrinsic theory of the sort suggested above would seem to offer a genuine alternative to the internal representation hypothesis as an intrinsic characterization of propositional attitudes.

The proposed alternative individuates propositional attitude states relationally, by their functional roles, but it does not analyse propositional attitudes as relations to internal entities. It is hard to tell whether such an account would violate Fodor's first adequacy condition. Fodor doesn't explicitly say in what sense propositional attitudes are to be analysed as relations. He presumably intends that propositional attitudes should be treated as relations between organisms

and entities of some sort, e.g. sentences; however, his arguments for the first adequacy condition -- that "believes" appears to be a two place predicate, that existential generalization applies to the syntactic objects of propositional attitude verbs, and that the fusion story is inadequate -- support only the claim that propositional attitude ascriptions should be analysed relationally. Construed as a condition on propositional attitude ascriptions, Fodor's first adequacy condition can be as easily satisfied by the functionalist alternative as by the internal representation account. The former theory is not wedded to any particular view of propositional attitude ascription, certainly not to the implausible fusion story. Construed as a condition on the attitudes themselves, the first adequacy condition cannot be satisfied by the functionalist alternative, since, on that account, propositional attitudes are not analysed as relations.

Fodor's second adequacy condition requires that a theory of propositional attitudes explain the parallelism between propositional attitude verbs and verbs of assertion: the things we can be said to believe, (want, hope, etc.) are the same things we can be said to assert. This parallelism is unsurprising on the functionalist theory, since propositional attitudes and natural language assertions are both mapped to the same intentional objects

(i.e. propositional contents).

According to Fodor's third adequacy condition, a "theory of propositional attitudes should account for their opacity."²² It is hard to know how to interpret this condition except as a constraint on a theory of propositional attitude ascriptions, since opacity is a fact about the sentences that ascribe propositional attitudes. Once again, the functionalist theory is not wedded to any particular account of propositional attitude ascription.

Fodor's fourth adequacy condition requires that the objects of propositional attitudes have logical form. The evidence that Fodor adduces in support of this requirement is of the following sort:

...John's belief that Mary and Bill are leaving is a conjunctive belief (cf. the logical form of "Mary and Bill are leaving"; John's belief that Alfred is a white swan is a singulary belief (cf. the logical form of "Alfred is a white swan"); and so on.... A theory of propositional attitudes should legitimize the ascription of form to the objects of propositional attitudes. In particular, it should explain why the form of a belief is identical to the logical form of the correspondent^[23] of a sentence which (opaquely) ascribes that belief. 24

Fodor gives no argument for the claim that the belief picked out by the belief ascription "John believes that Mary and Bill are leaving" has the form of a conjunction; he gives no argument for the claim that the belief has any

logical form at all. Propositional contents, of course, have logical form, and this property is exploited in formulating the theory's generalizations but it does not follow that the internal states to which propositional contents are mapped in the theory have logical form. The theory of propositional measurement proposes that we treat propositional contents as a scale for "measuring" propositional attitudes. We do not assume that temperature has all the properties of the scale used to represent it (we don't, for example, assume that something that is 40 degrees Fahrenheit has twice as much heat as something that is 20 degrees Fahrenheit). Of course, propositional attitudes themselves may have logical form; the propositional measurement theory is consistent with the hypothesis of a system of internal representation (cf. Field's construal of the propositional measurement theory, p.13). My point is simply that the assumption of logical form is not an adequacy condition on a theory of propositional attitudes.

To summarize what I have argued in this section: Fodor's first four adequacy conditions do not favour a theory which analyses propositional attitudes as relations between organisms and sentences in an internal language (e.g. the RTM) over an alternative proposal not committed to an internal system of representation. Whether empirical considerations (the fifth adequacy condition)

support the RTM will be addressed in the next section.

3. Arguments for the Language of Thought

In the appendix to Psychosemantics, entitled "Why There Still has to be a Language of Thought", Jerry Fodor argues that intentional realism is plausible only if we accept the language of thought thesis. Before turning to Fodor's arguments for this claim I would like to set out the three theses now in play. It is important to keep these theses separate when evaluating Fodor's arguments.

The first thesis, intentional realism, is the view that there really are beliefs and desires, or more precisely, that there are intentional states that (1) are semantically evaluable, (2) are causally efficacious in the production of behaviour, and (3) are truly described by the generalizations of commonsense psychology.

The second thesis is the Representational Theory of Mind, which claims that propositional attitudes are to be analysed as follows: for any organism O and attitude A toward the proposition p , there is a computational relation R and a mental representation MP such that MP means that p , and O has A to p if and only if O bears R to MP . The RTM further claims that mental processes are causal sequences of tokenings of mental representations.

The third thesis, the Language of Thought thesis (hereafter, LOT) is the view that mental states have constituent structure, or, more precisely, that mental states are relations to representations which are to be construed as sentence-like entities: they presuppose a language or symbol system.

A number of points need to be made about LOT. In the first place, it is not specifically a claim about propositional attitudes; it leaves open both the possibility that some propositional attitude tokenings are ~~not~~ relations to mental representations which have constituent structure, and the possibility that some tokenings of mental representations that have constituent structure have no plausible construal as propositional attitude tokenings. I shall use the term LOT_{pa} to refer to the thesis that propositional attitudes are to be construed as relations to (meaningful) token mental representations which have constituent structure.²⁵ LOT_{pa} , then, but not LOT, entails the RTM. I shall be concerned with whether Fodor can establish LOT_{pa} , since if LOT_{pa} is true, then so is the RTM.

Secondly, LOT_{pa} is, strictly speaking, a stronger thesis than the RTM. Steven Schiffer formulates a version of the RTM which is independent of LOT_{pa} .²⁶ On Schiffer's schematic formulation, there is a mechanism in the head corresponding to each propositional attitude

relation-type. For ease of exposition each of these mechanisms is just assumed to be a box; so there is a belief box, a desire box, etc. For each attitude tokening of a propositional content p , a symbol token that means that p is placed in the appropriate box. To believe that p , then, is just to have a symbol token that means that p in the belief_box. Propositional attitudes have their causal roles in virtue of whatever processing their associated symbol tokenings undergo in the boxes. On the assumption that the boxes are a simple way of talking about computational relations, Schiffer's story is just a picturesque formulation of the RTM.²⁷ It is not, however, committed to LOT_{pa} , because there is no requirement that the symbol tokens have constituent structure. They might be images, or indeed, as Fodor points out, rocks; all that is required is that they be semantically evaluable. According to LOT_{pa} , the symbols that go in the boxes have a sentence-like constituent structure. If a token has the content $p \& q$ then it has a constituent that means that p and a constituent that means that q . The symbol tokens, therefore, constitute an internal system of representation, i.e. a language.

I turn now to Fodor's arguments for LOT, to see if they establish LOT_{pa} . His first argument is methodological. He claims that the following is a plausible rule of nondemonstrative inference:

Principle P: Suppose there is a kind of event c_1 of which the normal effect is a kind of event e_1 ; and a kind of event c_2 of which the normal effect is a kind of event e_2 ; and a kind of event c_3 of which the normal effect is a complex event $e_1 \& e_2$. Viz.:

$c_1 \rightarrow e_1$

$c_2 \rightarrow e_2$

$c_3 \rightarrow e_1 \& e_2$

Then, ceteris paribus, it is reasonable to infer that c_3 is a complex event whose constituents include c_1 and c_2 . 28

Fodor takes principle P to be a special case of a general principle that requires us to prefer theories that minimize accidents. In defense of P he argues "if the etiology of events that are e_1 and e_2 does not somehow include the etiology of events that are e_1 and not e_2 , then it must be that there are two ways of producing e_1 events; and the convergence of these (ex hypothesi) distinct etiologies upon events of type e_1 is, thus far, unexplained."²⁹ To avoid an unexplained convergence of causes, Fodor concludes, we should infer that c_3 has c_1 and c_2 as constituents.

Principle P is then applied to psychological states. Behaviour, Fodor argues, is manifestly complex: it exhibits constituent structure. Verbal behaviour is the paradigm case -- verbal forms are put together from recurrent elements -- but even animal behaviour, bird song for example, exhibits constituent structure. Principle P, therefore, requires us to assume that the

causes of behaviour have constituent structure.

Fodor's argument for the claim that mental states have constituent structure rests on the plausibility of principle P. Principle P, however, is tendentious. Suppose that the normal effect of a particular bacterial infection is a sore throat, and the normal effect of a particular virus is a fever. Now suppose that a patient exhibits both a sore throat and a fever. Principle P would have us infer that the cause of the sore throat and fever is a complex condition consisting of both the bacterial infection and the virus! We are not normally tempted to assume a complex cause simply to minimize accidents. The sciences often postulate distinct etiologies converging on similar (identical) effects. "Accidents" are sometimes left to be explained by other theories; if we don't have the explanation in hand, then rather than assume a complex cause we sometimes just assume that an explanation of the "accident" will eventually be forthcoming, perhaps from a still to be developed micro-theory. Rather than appealing to counterexamples, though, it is more appropriate to ask of Fodor where Principle P comes from. Is it supposed to be an operative principle in the sciences? If so, what are some examples of its employment? Fodor gives no inductive support for the principle. It looks instead like an ad hoc device appealed to simply to get Fodor what he wants

-- that mental states have constituent structure. But if the principle has no independent support, and it seems to have none, then to appeal to it in the present context is question-begging. To grant Fodor's principle without some independent evidence of its plausibility is just to grant what is in dispute -- that mental states have constituent structure.

* * *

Fodor's second argument for LOT is based on current psychological research. He argues as follows:

...[Psycholinguists] say things like this: "When you understand an utterance of a sentence, what you do is construct a mental representation of the sentence that is being uttered. To a first approximation, such a representation is a parsing tree; and this parsing tree specifies the constituent structure of the sentence you're hearing, together with the categories to which its constituents belong. Parsing trees are constructed left to right, bottom to top, with restricted look ahead..." and so forth, depending on the details of the psycholinguist's story. Much the same sort of examples could be culled from the theory of vision (where mental operations are routinely identified with transformations of structural descriptions of scenes) or, indeed, from any other area of recent perceptual psychology. 30

Fodor points out that such theories appear to quantify over mental representations -- in the case of psycholinguistic theories over parsing trees -- and so

insofar as the theories are well-supported, we should recognize their commitment to mental representations.

I am prepared to grant that research in psycholinguistics and visual processing supports the claim that some mental processes involve token mental representations, but it should be noted that Fodor's argument, while it supports some versions of LOT, provides no evidence for LOT_{pa}. There is no natural construal of parsing trees, or other types of structural descriptions constructed in the course of language processing, as bearers of the content of propositional attitudes. Parsing trees are simply not representations of propositions or distal states of affairs to which an agent could plausibly be claimed to bear an attitude (believing, desiring, etc.).³¹ Parsing trees make explicit the grammatical structure of a sentence; but while the sentence may refer to distal states of affairs, the parsing tree does not. It simply provides an analysis of the sentence parsed. The rules of a speaker's grammar are more plausibly claimed to be interpretable as the contents of propositional attitudes (viz. what the speaker knows when she knows a language) than the structural descriptions that the grammar associates with the sentences of a language. Edward Stabler³² has convincingly argued, however, that the evidence that is typically cited in support of a particular grammar does

not support a construal of the grammar under which it is explicitly represented, over a construal under which the speaker is hardwired to compute the function which the grammar specifies.

The theory of vision is also claimed to provide support for LOT. Whether it supports LOT_{pa} will depend on whether the representations that are constructed in visual processing can be plausibly construed as token representations of propositional contents to which the subject bears an appropriate attitude (e.g. believing). Let's take David Marr's theory as an example of a promising theory of visual processing. The computational processes hypothesized by the theory are defined over token representations, so the theory is clearly committed to representations. But can the token representations be construed as bearers of propositional contents to which the subject bears an appropriate attitude? The answer isn't clear. The representational tokens which form the inputs to the stereopsis and directional selectivity processes, for example, are constructed from zero crossings, which, as pointed out in chapter 3, have no specifiable distal physical interpretation (they are not "physically meaningful"), so these representational tokens are not plausibly construed as bearers of propositional content. Whether the inputs and outputs to the other processes can be so interpreted is not clear. But a

number of considerations suggest that such an interpretation would be difficult.

First, the representational tokens postulated by Marr specify properties of surfaces -- intensity values in the case of the image, intensity changes and their geometrical distribution in the case of the primal sketch, and orientation and depth in the case of the 2.5-D sketch. The construction of these representational tokens precedes the decomposition of the scene into objects or otherwise meaningful regions. Marr says the following about the computational processes which take representational tokens as inputs and outputs:

Most early visual processes extract information about the visible surfaces directly, without particular regard to whether they happen to be part of a horse, or a man, or a tree. It is these surfaces -- their shape and disposition relative to the viewer -- and their intrinsic reflectances that need to be made explicit at this point in the processing, because the photons are reflected from these surfaces to form the image, and they are therefore what the photons are carrying information about. 33

They rely on information from the image...and the information they specify concerns the depth or surface orientation at arbitrary points in an image, rather than the depth or orientation associated with particular objects. 34

The content that the representational tokens have, therefore, concerns reflectance properties at arbitrary points in the image. If these representational tokens are

construed as bearers of propositional content to which the subject bears an attitude, then the subject bears an attitude toward such things as intensity changes in the image, discontinuities in surface orientation, etc. Nothing precludes the subject having propositional attitudes towards these contents, but they hardly look like central cases of beliefs and desires. The representational tokens postulated as the inputs and outputs to the visual processes are not interpreted in the theory as bearers of propositional contents to which the subject bears an attitude. Whether they can be so interpreted is an open question; but even if they can be they would confer no support on LOT_{pa} (or on the RTM), because they do not correspond to any independently ascribable attitudes.

There is a second, more general, reason for not taking the representational tokens postulated in Marr's theory to be likely candidates as bearers of propositional contents to which a subject bears an appropriate attitude. Linguistic entities are more plausible candidates for such a job than pictorial representations. The problem is that pictures don't really have determinate propositional content. Imagine someone holding up a picture and saying "This is the content of Jones' attitude". The appropriate response is "what is the content?". The point is perhaps best summed up in the old adage that "a picture is worth a

thousand words". Exactly what propositional content does a picture have? Furthermore, all versions of the LOT claim that the hypothesized mental representations form a language, that they have a sentence-like structure (i.e. a syntax). Perhaps the representations postulated in Marr's theory can be construed as quasi-linguistics entities, but this would need argument.

The above arguments are not intended to be decisive. I am not saying that the representations postulated in perceptual theories cannot be construed as bearers of propositional content to which an agent bears an appropriate attitude. But the case hasn't been made that they can. Merely citing the existence of representations to which perceptual theories are committed provides no support for LOT_{pa}.

One final point with respect to Marr's theory: the assumptions that reflect physical constraints and facilitate the recovery of shape information from properties of the image (e.g. the rigidity assumption) might more naturally be construed as representations of propositional contents to which the subject bears an attitude than the representational tokens which form the inputs and outputs to the computational processes. However, because these assumptions are not explicitly represented in the visual system, they provide no support for LOT_{pa}. Indeed, to the extent that they can be

construed as the contents of propositional attitudes, they are counterexamples to LOT_{pa} (and the RTM).

* * *

Fodor's third argument for LOT cites two properties that thoughts are alleged to have -- productivity and systematicity. It is then claimed that the best (only?) way to explain the productivity and systematicity of thought is by assuming that thoughts have constituent structure.

It is claimed that there is a potentially infinite set of, for example, belief-state types, since beliefs are type-individuated by their propositional contents and the set of propositions is infinite.³⁵ The productivity of the attitudes -- the fact that we are capable of thinking an infinite number of thoughts -- is best explained, Fodor argues, by the assumption that belief-states (desire-states, etc.) have a combinatorial structure. But Fodor does not press the productivity argument. Given that we are finite beings, it is only under idealization that belief-states are potentially infinite. Without idealization, the attitudes are not productive. The opponent of LOT, Fodor points out, can disarm the productivity argument simply by refusing to idealize. Such a refusal would seem to be ad hoc. The idealization

presupposed by the productivity argument seems innocuous: what prevents us from realizing an infinite number of belief-states are memory limitations and the eventual breakdown of our bodies, not any presumed limitations on the processes of belief formation itself. To refuse to idealize seems arbitrary. (We will see below that the opponent of LOT has a better response to the productivity argument.)

Fodor presents a second argument for LOT that does not depend on idealization. He develops the argument in a number of stages:

(1) Linguistic ability is systematic -- that is, "the ability to produce/understand some of the sentences [of a language] is intrinsically connected to the ability to produce/understand many of the others."³⁶ As an example Fodor cites the fact that speakers of English who know how to say that Mary loves John also know how to say that John loves Mary. The systematicity of linguistic ability is explained by the fact that natural languages have a combinatorial semantics -- the meaning of a complex expression is a function of the meaning of its parts. More concretely, our linguistic ability is said to be systematic because what we learn when we learn language is a set of rules. The same rule accounts for our ability to understand both "John loves Mary" and "Mary loves John".

(2) The sentences of natural language have a combinatorial

semantics because they have constituent structure; that is, they are built up out of recurrent words and phrases.

(3) Cognitive capacities are also systematic. Fodor argues as follows:

...cognitive capacities must be at least as systematic as linguistic capacities, since the function of language is to express thought. To understand a sentence is to grasp the thought that its utterance standardly conveys; so it wouldn't be possible that everyone who understands the sentence 'John loves Mary' also understands the sentence 'Mary loves John' if it weren't that everyone who can think the thought that John loves Mary can also think the thought that Mary loves John. You can't have it that language expresses thought and that language is systematic unless you also have it that thought is as systematic as language is. 37

(4) Cognitive capacity is systematic because thoughts have constituent structure. The idea that the systematicity of cognitive capacities implies the combinatorial structure of thoughts, Fodor claims, does not need defense. Given the systematicity of cognitive capacities he gets the combinatorial structure of thoughts "for free for want of an alternative account."³⁸

The systematicity argument, if correct, would support LOT_{pa}. If thoughts have combinatorial structure, then propositional attitudes have combinatorial structure. But does the argument support the claim that thoughts have combinatorial structure? It seems undeniable that thought is systematic. Anyone who can think the thought that John

loves Mary can think the thought that Mary loves John. So Fodor's argument is okay up to step (3). But is it clear that it is impossible for thought to be systematic without thoughts having constituent structure? Granted, propositional contents have constituent structure. But, on the account of propositional attitudes articulated in section 2, propositional attitudes are not analysed as relations to sentences or propositions; hence it does not follow from the fact that the propositional contents to which internal states are mapped have constituent structure that the internal states themselves have constituent structure.

In step (4) of the argument Fodor claims that he gets the combinatorial structure of thoughts "for free", for want of a better explanation of the systematicity of thought. But this clearly won't do. Both language and thought exhibit systematicity. While it is certainly possible that thought is systematic because thoughts have constituent structure, it is also possible that the systematicity of thought is merely exhibited or expressed, by exploiting something else which has constituent structure, viz. language. Thought is indeed systematic, but not necessarily because it has constituent structure. Fodor needs an independent argument for the latter claim. Thought ascriptions, of course, have constituent structure, but they are simply the means by which we

represent the systematicity (and productivity under idealization) of thought.

In a number of places Fodor offers a somewhat different argument for the claim that thoughts have constituent structure. He says

It seems inescapable -- and is, as far as I know, untendentious -- that the semantic properties of some beliefs (call them the complex ones) are inherited under some sort of combinatorial operations from the semantic properties of other beliefs (call them the simple ones). 39

Fodor considers a possible way of accommodating the above fact: some propositional attitudes might be analysed as complexes built out of simpler attitudes, without construing propositional attitudes as relations to mental representations which have constituent structure. On this proposal, the state of believing that P & Q is analysed as a complex state which has as constituents the state of believing that P and the state of believing that Q. While the proposal might work for conjunctive beliefs, it is, Fodor points out, inadequate as a general account:

Believing that P is not a constituent of, for example, believing that P or Q (or of believing that if P then Q...etc.); for it is perfectly possible to believe that P or Q (or that if P then Q) and not to believe that P. For similar reasons the required notion of constituency can't be defined over the causal roles of the attitudes, either. Thus, the causal role of believing that P is not a constituent of the causal role of believing

that P or Q since, for example, the effects of believing that it will snow in August are categorically different from -- and are not included among -- the effects of believing that either it will snow in August or it won't.

40

Clearly, the proposal to analyse propositional attitudes as complexes built out of simpler attitudes will not work. But there is no reason why the state of believing that P & Q (or the state of believing that $P \vee Q$, or the state of believing that if P then Q) must be analysed as a complex state. On the account proposed in section 2, what is complex is the propositional content to which the internal state of the organism is mapped, not the internal state itself. The propositional content P & Q is a complex content comprising the content that P and the content that Q, but these constituent contents are not themselves mapped to constituents of the internal state indexed by P & Q. The view that Fodor criticizes in the above quotation runs into trouble because it accepts the claim that some propositional attitude tokenings (e.g. a tokening of the belief that P & Q) are complex states, while also denying that propositional attitudes are relations, in particular, relations to mental representations. Once the former claim is granted, objects for which syntactic relations are defined (e.g. mental representations or natural language sentences) are the most plausible candidates for defining constituent

relations over (as Fodor points out above, defining the constituent relations over causal roles won't work). There is, however, no independent argument for the claim that some propositional attitude tokenings are complex states. Unless propositional attitudes are analysed as relations (i.e. unless a central tenet of the RTM is accepted as true) there is no clear sense to Fodor's supposedly "untendentious" claim that the semantic properties of some beliefs are inherited from the semantic properties of other beliefs.

Fodor's two arguments from the systematicity of thought both depend on the same fallacious reasoning. The propositional contents to which thoughts are mapped have constituent structure, and the sentences that ascribe thoughts have constituent structure, but a further argument is needed to establish that thoughts themselves have constituent structure.

A final point about productivity: the set of belief-state types is claimed to be infinite, because beliefs are individuated by their propositional contents and the set of propositions is infinite. According to Fodor, the fact that we are in principle capable of thinking an infinite number of thoughts is best (only) explained by the assumption that thoughts have a combinatorial structure. But other analyses of propositional attitudes are consistent with the

productivity of the attitudes. On the account sketched in section 2, since propositional attitudes are individuated by the propositional contents to which they are mapped, the set of possible propositional attitudes is, in principle, infinite. Of course, on both Fodor's and my accounts, propositional attitudes are internal states of the organism, and as such cannot in fact be infinite, since the organism's possible internal states are not infinite. The productivity of the attitudes per se is, therefore, not something that needs to be explained; it follows from the way that we represent propositional attitudes in natural language.⁴¹ On the account sketched in section 2 the attitudes are assumed to be productive because they can be mapped onto (and described by reference to) a system which is itself productive, i.e. propositions or public language sentences. To be sure, one explanation for why such a mapping is possible is that propositional attitudes are relations to sentences in an internal language. The alternative explanation sketched in section 2 would suggest that the representational system that we use to specify propositional attitudes -- the system of propositional contents -- has evolved with the properties it has, e.g. productivity, because it captures the salient functional properties of thought, including, for example, systematicity.

To summarize what I have argued in this section:
 Fodor's arguments do not establish LOT_{pa} -- the view that propositional attitudes are to be analysed as relations to mental representations which have constituent structure. The arguments that bear on LOT_{pa} either depend upon question-begging assumptions (principle P) or involve reading, without argument, properties possessed by propositional attitude ascriptions into propositional attitudes themselves. Fodor has made a case that some psychological theories are committed to the existence of mental representations, but since there is no plausible construal of the postulated representational tokens as the bearers of propositional content to which the subject bears an appropriate attitude, such theories do not provide support for LOT_{pa}, nor for the RTM.

4. Intentional Realism Without Explicit Representation

I have tried in this chapter to show that the RTM does not enjoy the support often claimed for it by its proponents. Fodor's five adequacy conditions on a theory of propositional attitudes do not favour the RTM over an alternative analysis of propositional attitudes not committed to the existence of explicit representation. Empirical considerations (e.g. the systematicity of

thought, the apparent commitment of some psychological theories to mental representations) provide no evidence for the RTM. Additionally, there exists a class of apparent counterexamples for which propositional attitude ascriptions seem justified, but which are not plausibly analysed as relations to internal representations. Furthermore, these apparent counterexamples threaten the RTM's prospects for vindicating folk psychological explanations of behaviour.

It is not my intention to argue that the RTM is false -- it may well be true. If it were "the only game in town", as its proponents claim, then counterexamples would not be decisive against it. But the only-game-in-town argument carries weight only if there is indeed no alternative in sight. I have sketched the outlines of an alternative; given that the RTM is itself neither well-supported nor (as a theory of propositional attitudes) particularly well-developed, a worked-out alternative should not be demanded. I shall conclude by considering to what extent the view sketched in section 2 might justify a realist interpretation of propositional attitude ascriptions.

According to the propositional measurement theory, contents can be assigned to internal states of the agent such that the structure of causal relations holding among the agent's internal states is reflected in the logical

relations holding among the associated propositional contents. The precise specification of how the causal relations are preserved in the mapping would be given by a representation theorem. The causal relations holding among the agent's internal states are then to be characterized independently of the propositional contents to which they are mapped by an intrinsic theory which specifies the functional roles of the agent's internal states.

The question I want to address here is this: to what extent could the propositional measurement theory support intentional realism, i.e. the view that there are internal states that (1) are semantically evaluable, (2) play a causal role in the production of behaviour, and (3) are truly described by the generalizations of commonsense psychology? The propositional measurement theory might seem to suggest an instrumentalist construal of propositional attitude ascriptions, since propositional contents, on the measurement theory, play essentially the role of measurement devices. Such a construal would be a mistake. There is no more reason to think that sentences of the form "x believes that p" cannot be true on the propositional measurement theory than there is to think that sentences of the form "x has a temperature of n degrees F." cannot be true. I shall argue that the propositional measurement theory supports a realist

construal of attitude ascriptions at least to the extent that the RTM does. Y

Pursuing the temperature analogy further, it should be noted that a realist construal of temperature ascriptions does not require that every property of the representation scheme be preserved in the mapping from physical objects to numbers. What must be preserved are certain ordinal relations defined on the integers, but not all arithmetic relations. As I pointed out earlier, we do not assume that an object that is forty degrees F. has twice as much mean kinetic energy as an object that is twenty degrees F.

Similarly, it is an implausibly strong requirement on a realist construal of propositional attitude ascriptions that every property of propositional space be preserved in the mapping to internal states. The question then is: what properties of propositional space do have to be preserved in the mapping to license realism about propositional attitudes? Minimally, what would have to be preserved is the following: (1) semantic evaluability, and (2) the pattern of relations among entities in the space. With respect to (2), the mapping must preserve the causal relations assumed (by folk psychology) to hold among propositional attitudes.⁴² More specifically, the causal relations among functional states, as characterized by the intrinsic theory, must be isomorphic to the relations

assumed by folk psychology to hold among propositional attitudes. A further requirement might plausibly be levied: that the intrinsic theory posit functional relations among internal states that would explain why propositional attitudes are appropriately individuated in terms of both a content and an attitude to that content (e.g. belief, desire). To meet this explanatory requirement the intrinsic theory would probably have to posit functional relations roughly coextensive with the various attitude-types. Satisfaction of the above conditions would justify the claim that the generalizations of folk psychology are true of the internal states postulated by the intrinsic theory; satisfaction of the third requirement would explain why those generalizations are true.

The RTM requires, in addition to the above conditions, that the postulated internal states have constituent structure. But, I have argued, this requirement has not been adequately supported. We have seen that constituent structure is not always preserved in the mapping. Dennett's chess-playing device is a case in point: a propositional attitude ascription seems justified but there is no internal state with the appropriate constituent structure. There is simply no reason to require that constituent structure be preserved in the mapping. The weaker requirements suggested above would

seem to be sufficient for vindicating propositional attitudes.

Of course, we have no guarantee that even the weaker requirements on intentional realism will be satisfied. In the event that no independent characterization of psychological states is forthcoming -- both the RTM and the functionalist alternative I have suggested fail, and folk psychology remains our only means of predicting and systematizing behaviour -- we might weaken the requirements on a realistic construal of attitude ascriptions still further. Absent any account of the underlying states we might simply say that folk psychology provides true characterizations of internal states, although we have no idea what makes these characterizations true. The requirements for realism with respect to propositional attitude ascriptions are, at this stage of cognitive theorizing, up for grabs. The strong requirements levied by Fodor are, at the very least, premature.

Notes

1. Fodor, J.A., Psychosemantics, Cambridge, Mass.: MIT Press, 1987, p.17.
2. Ibid., p.26.
3. Dennett, D.C., "A Cure for the Common Code," in Brainstorms, Vermont: Bradford Books, 1978, p.107.
4. Fodor admits that "the program actually operates on this principle ['get your queen out early']."
Psychosemantics, p.22.
5. Dennett denies that propositional attitude ascriptions carry commitments about the etiology of the device's behaviour. (see "Intentional Systems", in Brainstorms)
6. Ibid., pp.24-25.
7. Reprinted in RePresentations, Cambridge, Mass.: MIT Press, 1981.
8. Ibid., p.181.
9. Ibid., p.185.
10. Ibid., p.186.
11. Problems include indeterminacy of translation, and circularity if translation sets are explicated in terms of the intentions of speakers.
12. Ibid., p.197.
13. Churchland, Paul M., Scientific Realism and the Plasticity of Mind, Cambridge: Cambridge University Press, 1979, pp.100-107. The analogy is cited with approval in Dennett, D.C., "Beyond Belief," reprinted

- in The Intentional Stance, Cambridge, Mass.: MIT Press, 1987, and elaborated by Hartry Field in "Thought Without Content," unpublished manuscript.
14. Scientific Realism and the Plasticity of Mind, p.105.
 15. Field, Hartry. "Thought Without Content", unpublished manuscript, p.42.
 16. Ibid., pp.102-104. For a more perspicuous account see Field, Hartry, "Mental Representation" (postscript) reprinted in Block, Ned (ed.), Readings in the Philosophy of Psychology, Vol.2, Cambridge, Mass.: Harvard University Press, 1981. My explication draws on Field's account.
 17. It should be noted that Churchland does not endorse the propositional measurement theory as a way of doing psychology. He sees it simply as a means of purging propositional attitude ascription of an apparent commitment to relations between persons and abstract entities (in particular, propositions). But since he thinks (for reasons I will not go into here) that folk psychology is a "degenerating research program", he sees no place for propositional attitude ascription in serious psychology.
 18. Ibid.; p.114.
 19. For evidence that actual reasoning processes often violate logical strictures see Nisbett, R. and Ross, L., Human Inference: Strategies and Shortcomings of

Social Judgement, Englewood Cliffs, NJ: Prentice Hall, 1980; Tversky, A. and Kahneman, D., "Judgement Under Uncertainty: Heuristics and Biases," Science, 185, 1974; and Wason, P. and Johnson-Laird, P., Psychology of Reasoning: Structure and Content, London: B.T. Batsford, 1972.

20. For an detailed working out of a similar proposal see Brian Loar, Mind and Meaning, Cambridge: Cambridge U. Press, 1981. The view defended by Robert Stalnaker in Inquiry (Cambridge, Mass.: MIT Press, 1984) is in the same spirit. Whether sentences in the language of the belief-ascriber (Loar) or propositions (analysed either as sets of possible worlds (Stalnaker) or as structured quasi-linguistic entities) would be better "measures" of content is an important question, but my discussion of the propositional measurement theory will not presuppose any particular answer to it. The chosen measure of content must be fine-grained enough for use in characterizing the functional relations that hold among internal states, as specified by the intrinsic theory. Consequently, the issue of what entities to use as measures of content cannot be firmly resolved until we have an intrinsic theory in hand.
21. That Field does not construe the internal representation hypothesis to involve only a commitment to a structure-preserving mapping between propositions

and internal states is supported by the fact that the position that he articulates and defends in "Mental Representation" is a strong version of the internal representation thesis, viz. the RTM.

22. Fodor, "Propositional Attitudes," in RePresentations, p.181.

23. If "A believes that S" is a belief sentence then "S" is the correspondent of the sentence.

24. Ibid., p.185.

25. Fodor does not distinguish between LOT and the thesis I am calling LOT_{pa}, but I think it is important to do so. Fodor clearly wants to establish that the symbol tokens implicated in propositional attitude tokenings have constituent structure. The LOT is a vaguer thesis. It might be construed as the highly implausible strong claim that all mental states have constituent structure, or as the weaker thesis that some mental states have constituent structure. In any event, I shall be concerned only with Fodor's attempts to establish LOT_{pa}.

26. Although Fodor describes Schiffer's account as a version of the LOT, he points out that it is not committed to mental representation tokens having any constituent structure. Schiffer's formulation exactly captures the commitments of the RTM. Fodor is not always careful to distinguish the RTM and the LOT.

27. Fodor claims, incorrectly, that the Schiffer story is committed only to intentional realism. But intentional realism is not committed to the view that propositional attitudes are relations to symbol tokens. It claims only that propositional attitudes (1) are semantically evaluable, (2) play a causal role in the production of behaviour, and (3) are truly described by commonsense psychology.
28. Psychosemantics, p.141.
- 29 Ibid., p.142.
30. Ibid., pp.143-144.
31. Robert Matthews has argued this point in "The Alleged Evidence for Representationalism," forthcoming.
32. Stabler, E.P. Jr. "How are Grammars Represented?", The Behavioural and Brain Sciences, 3, 1983.
33. Marr, David. Vision, New York: Freeman, 1982, p.272.
34. Ibid., p.275.
35. Fodor doesn't actually give this argument. In "Why There Still Has to be a Language of Thought" he assumes, without argument that thoughts are potentially infinite. But there are a number of ways that one might arrive at the productivity of thoughts. One is by the argument I have given in the text -- that thoughts are individuated by reference to propositions, which themselves constitute an infinite set. A different argument bases the productivity of thoughts

on the productivity of natural language, arguing that since language expresses thought, thought must itself be productive.

36. Psychosemantics, p.149.
37. Ibid., p.151.
38. Ibid., p.151.
39. RePresentations, (Introduction), p.29. .
40. Pyschosemantics, pp. 166-167 (fn.5). Fodor gives a similar argument in RePresentations, pp.29-30.
41. As always, we have to be careful that properties of the measuring scale are not automatically read into the thing measured. It might be claimed that temperature is infinite because it is measured by the natural numbers.
42. These causal relations are represented as logical relations among propositions only under an idealization. See fn. 19 and related text p.21.

Bibliography

- Armstrong, D. A Materialist Theory of the Mind. London: Routledge and Kegan Paul, 1968.
- Block, N., ed. Readings in the Philosophy of Psychology, 2 vols. Cambridge, MA.: Harvard University Press, 1980.
- Burge, T. "Individualism and the Mental," Midwest Studies in Philosophy, IV, 1979, pp. 73-121.
- Burge, T. "Other Bodies," in A. Woodfield, ed., Thought and Object. Oxford: Oxford University Press, 1982.
- Burge, T. "Two Thought Experiments Reviewed," Notre Dame Journal of Formal Logic, 23, 1982, pp. 284-293.
- Burge, T. "Individualism and Psychology," Philosophical Review, XCV, no.1, 1986, pp. 3-46.
- Churchland, P.M. Scientific Realism and the Plasticity of Mind. Cambridge: Cambridge University Press, 1979.
- Dennett, D.C. "A Cure for the Common Code?" in Brainstorms. Cambridge, MA: MIT Press, 1978.
- Dennett, D.C. "Intentional Systems," in Brainstorms. Cambridge, MA: MIT Press, 1978.
- Dennett, D.C. "Beyond Belief," in The Intentional Stance. Cambridge, MA: MIT Press, 1987.
- Field, H. "Mental Representation," in Block, ed., Readings in the Philosophy of Psychology, Vol. 2, Cambridge, MA: Harvard University Press, 1980.
- Field, H. "Thought without Content," unpublished.
- Fodor, J.A. The Language of Thought. New York: Thomas Y. Crowell, 1975.
- Fodor, J.A. "Propositional Attitudes," The Monist, 64, no. 4, 1978. (Reprinted in RePresentations)
- Fodor, J.A. "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology," Behavioural and Brain Sciences, 3, 1980, pp. 63-110. (Reprinted in RePresentations)

- Fodor, J.A. RePresentations. Cambridge, MA: MIT Press, 1981.
- Fodor, J.A. "Cognitive Science and the Twin-Earth Problem," Notre Dame Journal of Formal Logic, 23, 1982, pp. 98-118.
- Fodor, J.A. "Reply to Brian Loar's 'Must Beliefs Be Sentences?'," in P. Asquith and T. Nichols, eds. Proceedings of the Philosophy of Science Association for 1982. East Lansing, Michigan, 1983.
- Fodor, J.A. "Observation Reconsidered," Philosophy of Science, 51, 1984, pp. 23-43.
- Fodor, J.A. "Banish Discontent," in J. Butterfield, ed., Language, Mind, and Logic. Cambridge: Cambridge University Press, 1986.
- Fodor, J.A. Psychosemantics: The Problem of Meaning in the Philosophy of Mind. Cambridge, MA: MIT Press, 1987.
- Fodor, J.A. "Is Content What 'That'-Clauses Pick Out? Reflections on Some Examples of Brian Loar's," unpublished.
- Kahneman, D. and Tversky, A. "Judgement Under Uncertainty: Heuristics and Biases," Science, 185, 1974, pp. 449-518.
- LePore, E. and Loewer, B. "Solipsistic Semantics," Midwest Studies in Philosophy, X, 1986, pp. 595-614.
- Loar, B. Mind and Meaning. Cambridge: Cambridge University Press, 1981.
- Loar, B. "Must Beliefs Be Sentences?," in P. Asquith and T. Nickles, eds., Proceedings of the Philosophy of Science Association for 1982, East Lansing, Michigan, 1983.
- Loar, B. "Social Content and Psychological Content," in D. Merrill and R. Grimm, eds., Content of Thought. Tucson: University of Arizona Press, forthcoming.
- Marr, D. Vision. New York: Freeman, 1982.

Marras, A. "The Churchlands on Methodological Solipsism and Computational Psychology," Philosophy of Science, 52, 1985, pp. 295-309.

Marras, A. "Critical Notice: Stephen Stich's From Folk Psychology to Cognitive Science: The Case Against Belief," Philosophy of Science, 54, 1987, pp. 115-127.

Matthews, R. "Troubles with Representationalism," Social Research, 51, 1984, pp. 1065-1097.

Matthews, R. "The Alleged Evidence for Representationalism," in S. Silvers, ed., Representationalism. Dordrecht: Reidel, forthcoming.

McClelland, J. and Rumelhart, D., eds., Parallel Distributed Processing: Explorations in the Microstructures of Cognition, 2 vols. Cambridge, MA: MIT Press, 1986.

Nisbett, R. and Ross, L. Human Inference: Strategies and Shortcomings. Englewood Cliffs: Prentice Hall, 1980.

Peacocke, C. "Explanation in Computational Psychology; Language, Perception, and Level 1.5," Mind and Language, 1, 1986, pp. 101-123.

Putnam, H. "The Meaning of Meaning," in Mind, Language, and Reality. Philosophical Papers, II. Cambridge: Cambridge University Press, 1975.

Pylyshyn, Z. "Computation and Cognition: Issues in the Foundation of Cognitive Science," Behavioural and Brain Sciences, 3, 1980, pp. 111-132.

Pylyshyn, Z. "Cognitive Representation and the Process-Architecture Distinction," Behavioural and Brain Sciences, 3, 1980, pp. 154-167.

Pylyshyn, Z. Computation and Cognition: Toward a Foundation for Cognitive Science. Cambridge, MA: MIT Press, 1984.

Ryle, G. The Concept of Mind. London: Hutchinson, 1949.

Skinner, B.F. "Behaviourism at Fifty," in T.W.Wann, ed., Behaviourism and Phenomenology: Contrasting Bases for Modern Psychology, Chicago: University of Chicago Press, 1964.

Smith, B. "Semantic Attribution and the Formality Condition," unpublished.

Stabler, E. "How are Grammars Represented?" Behavioural and Brain Sciences, 6, 1983, pp. 391-422.

Stalnaker, R. Inquiry. Cambridge, MA: MIT Press, 1984.

Sterelney, K. "Critical Notice: S.P. Stich, From Folk Psychology to Cognitive Science," Australasian Journal of Philosophy, 63, 1985, pp. 510-519.

Stich, S. From Folk Psychology to Cognitive Science: The Case Against Belief. Cambridge, MA: MIT Press, 1983.

Ullman, S. The Interpretation of Visual Motion. Cambridge, MA: MIT Press, 1979.

Wason, P. and Johnson-Laird, P. Psychology of Reasoning: Structure and Content. London: B.T. Batsford, 1972.

White, S. "Partial Character and the Language of Thought," Pacific Philosophical Quarterly, 63, 1982, pp. 347-365.

Woodfield, A. Thought and Object. Oxford: Oxford University Press, 1982.